

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 20:26:02 ; Search time 1873.74 Seconds

(without alignments)
13513.673 Million cell updates/sec

Title: US-09-805-919-3

Perfect score: 1210
Sequence: 1 tatattcatgtgaacacatg.....ttcgcaaaaaaaaaaaaaa 1210

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgc_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|-------|--------------------|--------|-------------|
| 1 | 58.8 | 4.9 | 611 | AX187051 |
| 2 | 55.6 | 4.6 | 33014 | CER06B9 |
| 3 | 55.4 | 4.6 | 7736 | AF310889 |
| 4 | 55.4 | 4.6 | 63807 | AC087299 |
| 5 | 55.4 | 4.6 | 112328 | AL390997 |
| 6 | 55.4 | 4.6 | 147891 | AL365191 |
| 7 | 55.4 | 4.6 | 183892 | AC019074 |
| 8 | 55.4 | 4.6 | 216680 | AC021211 |
| 9 | 55.2 | 4.6 | 19209 | AF274444 |
| 10 | 54 | 4.5 | 11662 | AX346802 |
| 11 | 53.8 | 4.4 | 234112 | PFMAL14P2 |
| 12 | 53.4 | 4.4 | 8952 | AX251200 |
| 13 | 53.4 | 4.4 | 17865 | AX345007 |
| 14 | 53.4 | 4.4 | 17865 | AX348307 |
| 15 | 53 | 4.4 | 198 | AX182091 |
| 16 | 53 | 4.4 | 169546 | AC004157 |
| 17 | 52.8 | 4.4 | 110000 | PFMAL14P1_2 |
| 18 | 52.4 | 4.3 | 318221 | PFMAL13P3 |
| 19 | 52.4 | 4.3 | 425 | AX284646 |
| 20 | 52.4 | 4.3 | 1462 | BC006018 |
| 21 | 52.4 | 4.3 | 8805 | AX345415 |
| 22 | 52.4 | 4.3 | 14920 | AX344738 |
| 23 | 52.4 | 4.3 | 46353 | AC100712 |
| 24 | 52.2 | 4.3 | 12592 | AX251796 |
| 25 | 52.2 | 4.3 | 14950 | AX346159 |
| 26 | 52.2 | 4.3 | 170941 | AC068027 |
| 27 | 52.2 | 4.3 | 172805 | AL360270 |
| 28 | 52.2 | 4.3 | 181936 | AC013387 |
| 29 | 52.2 | 4.3 | 184706 | AC013327 |
| 30 | 52.2 | 4.3 | 186462 | AL391727 |
| 31 | 52 | 4.3 | 6325 | DD1012088 |
| 32 | 51.8 | 4.3 | 77835 | PFMAL13P2_3 |
| 33 | 51.6 | 4.3 | 8342 | AX345403 |
| 34 | 51.6 | 4.3 | 205429 | AC005056 |
| 35 | 51.2 | 4.2 | 14553 | AE001391 |
| 36 | 51.2 | 4.2 | 56153 | AX251551 |
| 37 | 51.2 | 4.2 | 81358 | AC023330 |
| 38 | 51 | 4.2 | 975 | DD087514 |
| 39 | 51 | 4.2 | 12029 | AE001372 |
| 40 | 51 | 4.2 | 157373 | AC106404 |
| 41 | 50.8 | 4.2 | 5449 | AX346543 |
| 42 | 50.8 | 4.2 | 7657 | AX278022 |
| 43 | 50.8 | 4.2 | 7657 | AX346924 |
| 44 | 50.8 | 4.2 | 11209 | AX344729 |
| 45 | 50.8 | 4.2 | 13123 | AX281321 |

ALIGNMENTS

| RESULT | 1 | 611 bp | DNA | linear | PAT 06-AUG-2001 |
|---------------------|---|--------------------------------------|-----|--------|-----------------|
| AX187051/c | AX187051 | Sequence 2746 from Patent WO0124467. | | | |
| LOCUS | AX187051 | | | | |
| DEFINITION | AX187051 | | | | |
| ACCESSION | AX187051.1 | GI:15138495 | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| SOURCE | | | | | |
| ORGANISM | human. | | | | |
| | Homo sapiens | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 611) | | | | |
| AUTHORS | Schlegel, R., Deeds, J., Berger, A. and Zhao, X. | | | | |
| TITLE | Genes, compositions, kits, and methods for identification, assessment, prevention, and therapy of cervical cancer | | | | |
| JOURNAL | Patent: WO 0142467-A 2746 14-JUN-2001; | | | | |
| FEATURES | Millennium Predictive Medicine, Inc. (US) | | | | |
| location/Qualifiers | 1..611 | | | | |
| BASE COUNT | 379 a | 50 c | 2 g | 178 t | 2 others |
| ORIGIN | | | | | |


```

CDS
gene
    CQLPRATYPODDNLPQLPPTIGGCVAYAPPYCGIFPMADGSEKAGADSYAASMPM
    IPIPNVPSKDDGTAVVIPSAPPLSYESFSGDDGLTDEDEQFAKYVFNPLP
    VYNPTAPPE"
    join(27437..27621,27944..28049,28096..28270,28695..29366,
    29417..29562)
    /gene="R06B9.1"
    join(27437..27621,27944..28049,28096..28270,28695..29366,
    29417..29562)
    /gene="R06B9.1"
    /note="Similarity to Human dlhydroxyvitamin D3-induced
    protein"
    /codon_start=1
    /protein_id="CAB05786.1"
    /db_xref="GI:3878864"
    /db_xref="SPTREMBL:O17981"
    /translation="MPELELVIFDQPNVEYFPGQISGRVILSTTEYKARAVNIK
    ILGLATSWTDYDSVRVNDGNVRYDSESVHYSVHYLDQALLMACDSELSA
    GDVWPSPVPLPLNVPSPFEGKGYLKYVLTAVDRPRLDRAKRCITGEARLSPLI
    DLNAIPALPDIIDESENLGCCFFRKGYLELVNIPTKGFVGETVPMNIIHNS
    VPVEVAKAIIIOCKEIAFYRNGTIFRRDGSDDLMSGSSQOTKDYDKRPVITQDPMY
    TPGEHEKVELEFRLPSYTPRICRSPVITVEYVYVRYVETTSNGSAACEMPIIGT
    VPINRYLPPIPNRYPIGLPPRYANLTDVWVPCPSGSGTAVIPSAFPTOSMTIGV
    GGTLEKAEENEKPPAKTYFVFNPLPTYNTPAPSE"
    join(30328..30503,31116..31221,31318..31483,32023..32123,
    32186..32413)
    /gene="R06B9.2"
    join(30328..30503,31116..31221,31318..31483,32023..32123,
    32186..32413)
    /gene="R06B9.2"
    /note="Similarity to Human dlhydroxyvitamin D3-induced
    protein"
    /codon_start=1
    /protein_id="CAB05789.1"
    /db_xref="GI:3878867"
    /db_xref="SPTREMBL:O17984"
    /translation="MPELVIFDQPNVEYFPGQISGRVILSTTEYKARAVNIK
    KSHFTTQSEETKNSGSEETKTHVHTATAKHLDTAVPLMKRTDAAALDQK
    EMQWEPQLPQSVLPSPFEGKNGNIRYWRLEVSRSKFNIVDESELPDPLMTMP
    IARTPLDGFVAKNGCCCFRNGNSVSPDEQTVADIVHYSVNTYKDLAKILPQROSEN
    GNIIQGIILNGFERGIFHDGFELEGGLNGDHDGDIYFG"
BASE COUNT      10605 a      6288 c      5920 g      10201 t
ORIGIN

Query Match      4.68; Score 55.6; DB 3; Length 33014;
Best Local Similarity 52.18; Pred. NO. 0.047;
Matches 124; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY  939  tcttaagatttctgaagcatttaacagaagttgtatcttttccaagaaa 998
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10619 TTTTAAACATTTTAAACAAGATTTTGTGAAATTTTAAATTTTGTGAAAT 10560
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  999  tctaggaacaagtcacactgctgagcgaataactaagaatttcttcacgttag 1058
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10559 TCTTAGAATCTTTGCGACCTTTTGTGATTTTCTTAAACCTTTTCGCAAAATVATTG 10500
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1059  tccctatttattgatacttaccattatgtagtatttttccctaaat 1118
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10499 AATTTTITTTTGTGATTAATTTTCGAATATTTGGGGAATTTTGAACGAAAT 10440
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1119  ttataatttccatactcttgaagtaagtaagtaatttactcttgcacat 1176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10439 TCGAATTTTGTGAGATTTTGTAGGATTTTGTAGTGAATTCGAATTTTCGGTGAT 10382
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT  3
AF310889 AF310889 7736 bp DNA linear INV 06-MAR-2001
LOCUS Dictyostelium discoideum RacD (racd) gene, complete cds; and
ACCESSION AF310889
VERSION AF310889.1 GI:12007300
KEYWORDS

```

```

SOURCE
ORGANISM Dictyostelium discoideum.
Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE
AUTHORS Rivero,F., Dislich,H., Glockner,G. and Noegel,A.A.
TITLE The Dictyostelium discoideum family of Rho-related proteins
JOURNAL Nucleic Acids Res. 29 (5), 1068-1079 (2001)
MEDLINE 21127961
REFERENCE 2 (bases 1 to 7736)
AUTHORS Rivero,F., Dislich,H. and Noegel,A.A.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-2000) Institut fuer Biochemie I, Medizinische
Fakultaet, Universitaet zu Koeln, Joseph-Stelzmann-St. 52, Cologne
50931, Germany
FEATURES
source
    /organism="Dictyostelium discoideum"
    /strain="AX4"
    /db_xref="taxon:44689"
    join(<85..1007,1160..>4670)
    /product="unknown"
    join(856..1007,1160..4670)
    /note="5' region identified manually, therefore unsure;
    similarity to mouse SH3-domain containing adaptor protein
    (central part) and rat L-glutamine aminohydrolase
    (C-terminal part)"
    /codon_start=1
    /product="unknown"
    /protein_id="AAG45120.1"
    /db_xref="GI:12007301"
    /translation="MFKHQSNMLVYVFENIPANTKQDIAETSSSIGLVKGFGL
    NCEVCIFKSGSKAKIDQIVQVOMDKQGVKWSHLSEITSNKAVLFEYEYTS
    ELSLIPHOVIDIISKDGGMGESGKQNNGVSPFVQEPEDDQASAEAT
    KEERISKPIQSVGRILGNISGMDLKLKRYKATYATDIDEPANNNNAERS
    NSSSSSSSSNNISSGEFAVYKSGAVNQKQPKVILNNQSTILGCVPAIQTIP
    SSNTNTTTNTTPAVOPLQKPKOTNEILNNGSLPATTSKSOSTKESSGNGN
    KITSLEKPLKTKHTAISALGSHSSSSSGDKNNSGKTAKVYDFESKEPMELN
    LKGDIIIVLAKSGGVMQGINOSTQATQFWNTVEELKEERLDPSPSSSNKQPLK
    IYSPPEKSSSSSQPOVEAKLQSLDTPKLETKARAKAKRPPSRVASYLISDD
    EKAREKTKQSPFVSHVYSQKELSMYDNNDGDEHLSDDQSTPSPISRSLSSTP
    QQQQSGSVKPIERPSKPPPKRTAAMNSINNINPPPSKKPPPPKSLKSGSTPD
    GYRPPSPSSSTTPITTSGETDICKPTQKRVAPRQQQMDLPSSLVKPPPLK
    PTKPRALIQNSSSTAPADTPPSHEVNFKANLKAKPSPTGNSQIPIDVSSPNAT
    TPQPTPPLKRVAKPVVQQQPSNENIENTQPKLPSPASPTSSSPITDITSTTK
    PPARPPLKSAALAAPSPSPNEPATTPPTLAKKVALRNGPSAPSPSQOQP
    LPSTHRNLSPNTNNEKPPPTKPOLIKPLPQNDICVESHCKEIOERDSS
    QETTTFRKNYPSLIDSEFVAVCSNDQGFMGDDOLNHCYPMROYIYFYSIICE
    EIGLDEVSKFISQARPDSYDNOKPLNDOKKSHNPETLAGOLITANLSVNIOMVLE
    STQFINIICQELVNCSSVSCDMVSYLSOKSPTSHELIVGHLKSSGLIKQPDVLDPEY
    QINSIQNSQVSVLANTLNDVCFPSDDIKLAPNNIITSKTDLRICNSTPITSSI
    FNNIDINNSKIIPITISDSGVLMITIIIPGLMGVSIISPNNNNNNNNNNNNFHHNP
    NNHIEFKRLCSLIDQ"
    join(5116..5180,5293..5333,5523..5590,5805..5914,
    6023..6634)
    /gene="racD"
    /note="corresponds to EST clones SSR614, SSG550, SSJ204
    and SSK792 from the Japanese cDNA sequencing project; cDNA
    (partial) already described in L11594."
    /product="RacD"
    5116..6634
    /gene="racD"
    join(5163..5180,5293..5333,5523..5590,5805..5914,
    6023..6550)
    /gene="racD"
    /note="Rho-GTPase; lacks prenylation motif characteristic
    of Rho proteins"
    /codon_start=1
    /product="RacD"
    /protein_id="AAG45121.1"
    /db_xref="GI:12007302"
    /translation="MATGIKTKYKYVYVGDGAVGKTSLLIITTKAFKDVYPTFEDN
    FNCLEMIDKPNVLVMDTRAGQDIDYDNKLPSPQTDVITICYSYVAKSDSNITKRW
    LPEINQINQGTPIILVGTIKDLREKRTLSIQSESKQEPVSHDEGVALAKETGAQVF

```



```

* 18404 19087: contig of 684 bp in length
* 19088 19187: gap of 100 bp
* 19188 19882: contig of 695 bp in length
* 19883 19982: gap of 100 bp
* 19983 20676: contig of 694 bp in length
* 20677 20776: gap of 100 bp
* 20777 21473: contig of 697 bp in length
* 21474 21573: gap of 100 bp
* 21574 22282: contig of 709 bp in length
* 22283 22382: gap of 100 bp
* 22383 23092: contig of 710 bp in length
* 23093 23192: gap of 100 bp
* 23193 23872: contig of 680 bp in length
* 23873 23972: gap of 100 bp
* 23973 24731: contig of 759 bp in length
* 24732 24831: gap of 100 bp
* 24832 25530: contig of 699 bp in length
* 25531 25630: gap of 100 bp
* 25631 26333: contig of 703 bp in length
* 26334 26433: gap of 100 bp
* 26434 27122: contig of 689 bp in length
* 27123 27222: gap of 100 bp
* 27223 27894: contig of 672 bp in length
* 27895 27994: gap of 100 bp
* 27995 28670: contig of 676 bp in length
* 28671 28770: gap of 100 bp
* 28771 29478: contig of 708 bp in length
* 29479 29578: gap of 100 bp
* 29579 30282: contig of 704 bp in length
* 30283 30382: gap of 100 bp
* 30383 31081: contig of 699 bp in length
* 31082 31181: gap of 100 bp
* 31182 31853: contig of 672 bp in length
* 31854 31953: gap of 100 bp
* 31954 32660: contig of 707 bp in length
* 32661 32760: gap of 100 bp
* 32761 33474: contig of 714 bp in length
* 33475 33574: gap of 100 bp
* 33575 34291: contig of 717 bp in length
* 34292 34391: gap of 100 bp
* 34392 35072: contig of 681 bp in length
* 35073 35172: gap of 100 bp
* 35173 35857: contig of 685 bp in length
* 35858 35957: gap of 100 bp
* 35958 36627: contig of 670 bp in length
* 36628 36727: gap of 100 bp
* 36728 37438: contig of 711 bp in length
* 37439 37538: gap of 100 bp
* 37539 38248: contig of 710 bp in length
* 38249 38348: gap of 100 bp
* 38349 39043: contig of 695 bp in length
* 39044 39143: gap of 100 bp
* 39144 39842: contig of 699 bp in length
* 39843 39942: gap of 100 bp
* 39943 40657: contig of 715 bp in length
* 40658 40757: gap of 100 bp
* 40758 41433: contig of 676 bp in length
* 41434 41533: gap of 100 bp
* 41534 42176: contig of 643 bp in length
* 42177 42276: gap of 100 bp
* 42277 42985: contig of 709 bp in length
* 42986 43085: gap of 100 bp
* 43086 43793: contig of 708 bp in length
* 43794 43893: gap of 100 bp
* 43894 44609: contig of 716 bp in length
* 44610 44709: gap of 100 bp
* 44710 45405: contig of 696 bp in length
* 45406 45505: gap of 100 bp
* 45506 46211: contig of 706 bp in length
* 46212 46311: gap of 100 bp
* 46312 46988: contig of 677 bp in length
* 46989 47088: gap of 100 bp
* 47089 47802: contig of 714 bp in length

```

```

* 47803 47902: gap of 100 bp
* 47903 48614: contig of 712 bp in length
* 48615 48714: gap of 100 bp
* 48715 49436: contig of 722 bp in length
* 49437 49536: gap of 100 bp
* 49537 50219: contig of 683 bp in length
* 50220 50319: gap of 100 bp
* 50320 50991: contig of 672 bp in length
* 50992 51091: gap of 100 bp
* 51092 51777: contig of 686 bp in length
* 51778 51877: gap of 100 bp
* 51878 52599: contig of 722 bp in length
* 52600 52699: gap of 100 bp
* 52700 53420: contig of 721 bp in length
* 53421 53520: gap of 100 bp
* 53521 54215: contig of 695 bp in length
* 54216 54315: gap of 100 bp
* 54316 55020: contig of 705 bp in length
* 55021 55120: gap of 100 bp

```

```

Query Match 4.58; Score 55.4; DB 2; Length 63807;
Best Local Similarity 49.8%; Pred. No. 0.051;
Matches 140; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```

```

QY 928 atgaacgaactgttaagattttgtctagtcgaagctatttaacagaatttgtattt 987
Db 34594 AAGAAATGAGACCTGAAATTTGCTAGCAGACAGGTAAGTATGTAATAATTTATA 34653
QY 988 ttccagaataatcaggacaaggtcaactctgtctgcgatttaactaaggattttct 1047
Db 34654 ATTCTGAGTATATATATTTATGTCGCTGTGTAATAAGATTAATTTATTTCTT 34713
QY 1048 ttccagtttagctctgatttttttttttttttttttttttttttttttttttttt 1107
Db 34714 TTTCAGATGTTCTATAGATAGATATAGATATAGATATAGATATAGATATAGAT 34773
QY 1108 ttccataaatttataattttccataattcttcttaatttgaatgaatgaatgaat 1167
Db 34774 ATCTGAAACTTTATATATATTTATGCTATGCTATGCTATGCTATGCTATGCTAT 34833
QY 1168 tctgtcaataatagaacaagacattgcacaaaaaataaaaaa 1208
Db 34834 TTTCAAATATATAGATCATCATCTGCAACAGATTAATTTAA 34874

```

```

RESULT 5
AL390997/c 112328 bp DNA linear PRI 06-DEC-2001
LOCUS Human DNA sequence from clone RP11-1596 on chromosome Xq21.33-23,
DEFINITION complete sequence.
ACCESSION AL390997
VERSION AL390997.10 GI:10432544
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 112328)
REFERENCE 1
AUTHORS Howden, P.
TITLE Direct Substitution
JOURNAL Submitted (06-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humangery@sanger.ac.uk
On Sep 28, 2000 this sequence version replaced gi:10303411.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP, Information

```

COMMENT

on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/Celegans/wormpep> This sequence
 was generated from part of bacterial clone contigs of human
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChX>

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP11-1596 is from
 the library RP11-11.1 constructed by the group of Pieter de Jong.
 For further details see
<http://www.chori.org/bacpac/home.htm>

VECTOR: PBAC3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-1596. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true right end of clone RP11-1596 is at 112328 in this
 sequence. The true right end of clone RP13-13024 is at 100 in this
 sequence. The true right end of clone RP11-115W20 is at 27343 in
 this sequence.

FEATURES

source

```

1..112328
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q21.33-23"
/clone_lib="RP11-1596"
/clone_lib="RP11-11.1"
24..4692
/feature="LIM1 repeat: matches 1390..3476 of consensus"
4704..5475
/feature="LIMC3 repeat: matches 6280..7003 of consensus"
3489..3609
/feature="MER5A repeat: matches 1..121 of consensus"
8421..8505
/feature="MIR repeat: matches 164..244 of consensus"
10380..11089
/feature="MER21B repeat: matches 37..713 of consensus"
MER21B repeat: matches 37..713 of consensus
11416..11644
/feature="LIMC repeat: matches 1643..1875 of consensus"
11474..11756
/feature="HAL1 repeat: matches 625..912 of consensus"
12125..12487
/feature="LIM4 repeat: matches 2148..2523 of consensus"
12581..12850
/feature="LIM4 repeat: matches 2857..3140 of consensus"
12911..12958
/feature="MST-INTERNAL repeat: matches 1586..1635 of consensus"
MST-INTERNAL repeat: matches 1586..1635 of consensus
12966..13057
/feature="MST-INTERNAL repeat: matches 1586..1635 of consensus"
13112..13459
/feature="MST-INTERNAL repeat: matches 1233..1587 of consensus"
MST-INTERNAL repeat: matches 1233..1587 of consensus
13463..13575
/feature="MST-INTERNAL repeat: matches 1233..1587 of consensus"
13650..14015
/feature="MST-INTERNAL repeat: matches 1233..1587 of consensus"
14146..14224
/feature="MST-INTERNAL repeat: matches 1233..1587 of consensus"
14225..14537
/feature="MST-INTERNAL repeat: matches 1233..1587 of consensus"
14538..15365
/feature="MST-INTERNAL repeat: matches 1233..1587 of consensus"
15384..17055
/feature="LIM1 repeat: matches 1390..269 of consensus"

```

```

repeat_region 17055..18985
/feature="L1 repeat: matches 2851..4751 of consensus"
18982..19698
/feature="LIM2 repeat: matches 4957..5674 of consensus"
19716..19775
/feature="L1 repeat: matches 3926..3986 of consensus"
19786..20381
/feature="LIM2 repeat: matches 5681..6308 of consensus"
20484..20987
/feature="LIM2 repeat: matches 2802..3311 of consensus"
20988..21046
/feature="MIR repeat: matches 1..80 of consensus"
21047..23080
/feature="LIM2 repeat: matches 3311..5305 of consensus"
21489..21554
/feature="LIM2 repeat: matches 3311..5305 of consensus"
23108..26375
/feature="LIM2 repeat: matches 2872..6145 of consensus"
23110..26375
/feature="LIM2 repeat: matches 2872..6145 of consensus"
26376..27185
/feature="LIM2 repeat: matches 5302..6131 of consensus"
27279..27880
/feature="L1 repeat: matches 3729..4348 of consensus"
28059..28140
/feature="LIM3 repeat: matches 6109..6180 of consensus"
28173..29491
/feature="LIM3 repeat: matches 4812..6156 of consensus"
32110..32514
/feature="LIM4 repeat: matches 6513..6938 of consensus"
32573..32807
/feature="LIM4 repeat: matches 7026..7288 of consensus"
33293..34392
/feature="LIM4 repeat: matches 5025..6141 of consensus"
LIM4 repeat: matches 5025..6141 of consensus
34555..36345
/feature="LIM2 repeat: matches 4140..6155 of consensus"
36579..36893
/feature="LIM2 repeat: matches 4140..6155 of consensus"
37365..37532
/feature="MIR repeat: matches 180..510 of consensus"
38710..38847
/feature="LIM repeat: matches 5677..5802 of consensus"
39170..39353
/feature="MIR repeat: matches 109..237 of consensus"
40495..40572
/feature="L2 repeat: matches 3..210 of consensus"
42209..42346
/feature="MIR repeat: matches 2620..2710 of consensus"
42362..42804
/feature="MIR repeat: matches 1..139 of consensus"
42805..42866
/feature="LIM6 repeat: matches 5709..6157 of consensus"
43090..43703
/feature="LIM6 repeat: matches 123..186 of consensus"
44064..44723
/feature="LIM6 repeat: matches 2156..2749 of consensus"
44724..44771
/feature="LIM6 repeat: matches 5498..6160 of consensus"
44772..44908
/feature="LIM6 repeat: matches 5510..6156 of consensus"
44909..44912
/feature="LIM6 repeat: matches 256..303 of consensus"
44913..44916
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44917..44919
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44920..44922
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44923..44925
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44926..44928
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44929..44931
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44932..44934
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44935..44937
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44938..44940
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44941..44943
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44944..44946
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44947..44949
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44950..44952
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44953..44955
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44956..44958
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44959..44961
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44962..44964
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44965..44967
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44968..44970
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44971..44973
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44974..44976
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44977..44979
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44980..44982
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44983..44985
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44986..44988
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44989..44991
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44992..44994
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44995..44997
/feature="LIM6 repeat: matches 5391..5527 of consensus"
44998..44999
/feature="LIM6 repeat: matches 5391..5527 of consensus"

```



```

misc_feature      clone_end:T7
                  vector_side:left"
                  4197. .6285
                  /note="assembly_fragment:01135
                  fragment_chain:1"
misc_feature      6386. .14509
                  /note="assembly_fragment:01247
                  fragment_chain:2"
misc_feature      14610. 17060
                  /note="assembly_fragment:00977
                  fragment_chain:2"
misc_feature      17161. .21976
                  /note="assembly_fragment:01246
                  fragment_chain:2"
misc_feature      22077. .30708
                  /note="assembly_fragment:00221
                  fragment_chain:2"
misc_feature      30809. .46192
                  /note="assembly_fragment:01429
                  fragment_chain:3"
misc_feature      46293. .89106
                  /note="assembly_fragment:00855
                  fragment_chain:3"
misc_feature      89207. .57724
                  /note="assembly_fragment:00340"
                  97825. .103968
                  /note="assembly_fragment:00436"
                  104069. 106312
                  /note="assembly_fragment:00472"
                  106413. 110483
                  /note="assembly_fragment:00544"
                  110584. 117466
                  /note="assembly_fragment:00657"
                  117567. .127508
                  /note="assembly_fragment:00812"
                  127609. .130184
                  /note="assembly_fragment:01259
                  fragment_chain:4"
                  130285. 132386
                  /note="assembly_fragment:00064
                  fragment_chain:4"
                  132487. .141197
                  /note="assembly_fragment:00851
                  fragment_chain:4"
                  141298. .147891
                  /note="assembly_fragment:00995
                  fragment_chain:4
                  clone_end:SP6
                  vector_side:right"
misc_feature      26405 c 27064 g 48391 t 1712 others

```

| | Query Match | 4.6% | Score 55.4 | DB 2 | Length 14781 |
|-----------|---|----------------|----------------|----------------|--------------|
| | Best Local Similarity 49.8% | | Pred. No. 0.05 | Mismatches 141 | Indels 0 |
| | Matches 140 | Conservative 0 | | Gaps 0 | |
| QY 928 | atgaacgaacgttgaagattttgtcgtacgaacatatttaacagaagtttgcattt | 987 | | | |
| Db 108584 | AAGATATGGACCTGATTTTGCTGACACACAGGGTACTATAGTCAAAATATATTATA | 108525 | | | |
| QY 988 | tttcagaaaatctcaggaacaagtcacccgtgtcgtcgattatactcagaatttct | 1047 | | | |
| Db 108524 | ATTCCTTAGCTATATATATTTTATGTGGCGCTGTGTATAATATGATTAACGTTAATTTTCCTT | 108465 | | | |
| QY 1048 | ttccagttttagctcgtatttatttgcatacttctaactttgatttgcataagattt | 1107 | | | |
| Db 108464 | TTTCAGATTTGTTTCATATAGATGATATATGAAATGCTACTGATTTGTGAGTGATATTGTG | 108405 | | | |
| QY 1108 | tttccttaaaatttatataatttctcctaattcttgaagtaattgaatgataattgtactt | 1167 | | | |
| Db 108404 | ATCTCGAAACTTATATATTATTATCACTTGTATATAGTTTGTGGAGATATACAGGTT | 108345 | | | |

QY 1168 tctgtcattatagtaacaaagcatctgcgaataaaaaaaaaa 1208
| | | | | | | | | | | | | | |
Db 108344 TTTTCAAATATTAGATCATCATCTGCAACGATTAATTTA 108304

RESULT 7
AC019074/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC019074 18392 bp DNA linear HTG 07-JUL-20000
Homo sapiens chromosome X clone RP11-305M20, WORKING DRAFT
SEQUENCE, 12 unordered pieces.
AC019074
AC019074.3 GI:7923976
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 183892)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 183892)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 18, 2000 this sequence version replaced gi:7022215.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NHO3035M20
----- Summary Statistics -----
Sequencing vector: M13: 87%
Sequencing vector: plasmid: 13%
Chemistry: Dye-terminator ET, 87% of reads
Chemistry: Dye-terminator Big Dye, 13% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 178146 bases at least Q40
Consensus quality: 179992 bases at least Q30
Consensus quality: 180917 bases at least Q20
Insert size: 189000; agarose-fp
Insert size: 182792; sum-of-contigs
Quality coverage: 4.59 in Q20 bases; agarose-fp
Quality coverage: 4.78 in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
1258 1157: contig of 1257 bp in length
1358 3552: contig of 2295 bp in length
3653 3752: gap of unknown length
3753 10739: contig of 6987 bp in length
10740 10839: gap of unknown length
10840 10855: contig of 816 bp in length
18856 19055: gap of unknown length
19056 29630: contig of 10375 bp in length
29631 29530: gap of unknown length
44571 45771: contig of 15741 bp in length
45772 45371: gap of unknown length
45372 59160: contig of 13789 bp in length
59161 59260: gap of unknown length
59261 78447: contig of 18987 bp in length
78248 78347: gap of unknown length

```



```

* 18558 19722: contig of 1165 bp in length
* 19723 19822: gap of unknown length
* 19823 21952: contig of 2130 bp in length
* 21953 22052: gap of unknown length
* 22053 23772: contig of 1720 bp in length
* 23773 23872: gap of unknown length
* 23873 25143 25242: contig of 1270 bp in length
* 25143 25242: gap of unknown length
* 25243 26773 26872: contig of 1530 bp in length
* 26773 26872: gap of unknown length
* 26873 28395 28494: gap of 1522 bp in length
* 28395 28495 29921: gap of unknown length
* 29921 30021: contig of 1427 bp in length
* 30021 31831 31930: gap of unknown length
* 31831 31931 33274: contig of 1809 bp in length
* 33274 33373: gap of unknown length
* 33373 35937: gap of 1343 bp in length
* 35937 36037: contig of 2564 bp in length
* 36037 36038 38479 38578: gap of unknown length
* 36038 38479 38578: contig of 2441 bp in length
* 38479 38578: gap of unknown length
* 38579 41010: contig of 2432 bp in length
* 41010 41101: gap of unknown length
* 41101 43732: contig of 2622 bp in length
* 43732 43833: gap of unknown length
* 43833 45367: contig of 1535 bp in length
* 45367 45467: gap of unknown length
* 45467 47581 47680: contig of 2113 bp in length
* 47581 47681 49453: gap of unknown length
* 49453 49454 49553: contig of 1773 bp in length
* 49553 49554 51434: gap of unknown length
* 51434 51535 51536: contig of 1881 bp in length
* 51535 51536 55538: gap of unknown length
* 55538 55539 59217: contig of 4004 bp in length
* 59217 59218 59317: gap of unknown length
* 59317 59318 63649: contig of 3579 bp in length
* 63649 63650 63749: gap of unknown length
* 63749 63750 68922: contig of 4332 bp in length
* 68922 68923 69023: gap of unknown length
* 69023 73759 73758: contig of 5173 bp in length
* 73759 73758 73858: gap of unknown length
* 73858 73859 79377: contig of 4736 bp in length
* 79377 79378 79477: gap of unknown length
* 79477 79478 86657: contig of 5518 bp in length
* 86657 86658 86757: gap of unknown length
* 86757 86758 95803: contig of 7180 bp in length
* 95803 95804 95903: gap of unknown length
* 95903 95904 107023: contig of 9046 bp in length
* 107023 107024 107123: gap of unknown length
* 107123 107124 119084: contig of 11120 bp in length
* 119084 119085 119184: gap of unknown length
* 119184 119185 133685: contig of 11961 bp in length
* 133685 133686 133785: gap of unknown length
* 133785 133786 144764: contig of 14501 bp in length
* 144764 144765 144864: gap of unknown length
* 144864 144865 156086: contig of 10979 bp in length
* 156086 156087 156186: gap of unknown length
* 156186 156187 170774: contig of 11222 bp in length
* 170774 170775 170874: gap of unknown length
* 170874 170875 192194: contig of 14588 bp in length
* 192194 192195 192294: gap of unknown length
* 192294 192295 216680: contig of 21330 bp in length
* 216680 216681 24386: gap of unknown length

```

FEATURES

Source

```

1..216680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-407E4"
BASE COUNT 67337 a 38520 c 39310 g 66933 t 4580 others
ORIGIN

```

Query Match 4.6%; Score 55.4; DB 2; Length 216680;
 Best Local Similarity 49.8%; Pred. No. 0.049;
 Matches 140; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```

QY 928 atgaacgaactgttaagattttgtcgtcagcgaactctttacagaagttgtgtatt 987
DB 50223 AAGAAATGAGACCTGAAATTTGCTAGCACACAGGTAACCTATGATGCAAAAATTTTAA 50282
QY 988 ttccagaanaatcaggacaaggtcaacccgtcgtcgagtaataactagattttct 1047
DB 50283 ATCTCTAGGATATATATTTATGTCGCTGTGTAANAAGATACGTTATTTCTT 50342
QY 1048 ttccagtttagctcgtatattttatgtatattcttccattgtattgtattttt 1107
DB 50343 TTTCAGATGTCATTAATGATATAGAAATGCTACTGATTTGTAGTGATTTGCT 50402
QY 1108 ttccctaaatttataatttccctaatcttctgtaagtaattgagattgtactt 1167
DB 50403 ATCTGAAACTTATATATTTATATACACTTCTATATATTTTGTGAGATATTCAGGTT 50462
QY 1168 tctgtcaataataagaacaagacattcgcaaaaaa 1208
DB 50463 TTTCATATATATAGATCATCATCTGCAACGATTAATTAA 50503

```

RESULT 9
 LOCUS AF274444/c 19209 bp DNA linear BCT 11-MAR-2001
 DEFINITION

Carsonella ruddii ribosomal protein L11 (rpl11), ribosomal protein L1 (rpl1), ribosomal protein L10 (rpl10), ribosomal protein L7/L12 (rpl7/L12), beta subunit of RNA polymerase (rpoB), beta-prime subunit of RNA polymerase (rpoC), ribosomal protein S12 (rps12), ribosomal protein S7 (rps7), elongation factor G (efuA), elongation factor Tu (tuif), ribosomal protein S10 (rps10), ribosomal protein L3 (rpl3), ribosomal protein L4 (rpl4), ribosomal protein L2 (rpl2), ribosomal protein S19 (rps19), ribosomal protein L22 (rpl22), ribosomal protein S3 (rps3), ribosomal protein L16 (rpl16), ribosomal protein S17 (rps17), ribosomal protein L14 (rpl14), ribosomal protein L5 (rpl5), ribosomal protein S14 (rps14), and ribosomal protein S8 (rps8) genes, complete cds; and ribosomal protein L6 (rpl6) gene, partial cds.

ACCESSION AF274444
 VERSION AF274444.1 GI:13272287
 KEYWORDS

SOURCE

ORGANISM *Candidatus Carsonella ruddii*.

REFERENCE 1 (bases 1 to 19209) Bacteria; Proteobacteria; gamma subdivision; *Candidatus Carsonella*.

AUTHORS Clark, M.A., Baumann, L., Thao, M.L., Moran, N.A. and Baumann, P.
 TITLE Degenerative Minimization in the Genome of a Psyllid Endosymbiont
 JOURNAL J. Bacteriol. 183 (6), 1853-1861 (2001)

REFERENCE

AUTHORS Clark, M.A., Baumann, L., Thao, M.L., Moran, N.A. and Baumann, P.
 TITLE Direct Submission
 JOURNAL Submitted (01-JUN-2000) Microbiology Section, University of California Davis, One Shields Avenue, Davis, CA 95616-8665, USA

FEATURES

source Location/Qualifiers
 1..19209

gene /organism="Candidatus Carsonella ruddii"
 /specific_host="Pachysylla venusta"
 /db_xref="taxon:114166"
 221..649
 CDS /gene="rpl11"
 221..649
 /gene="rpl11"
 /codon_start=1
 /transl_table=11
 /product="ribosomal protein L11"
 /protein_id="AAK17071.1"
 /db_xref="GI:13272288"

```

gene
/translation="MKILKFKLILKPKAPSSGIGPILGOYGINLIDECNKNFNL
TKFICLEKINVKIAIFDNDNSYOILINTMSLSYIEKKNILISFSKPKPGYNQDTNINLN
DVEIISYFERNIEKKKLNLSRKMMLGTLISMGYENKKYK"
CDS
630..1205
/gene="rpL1"
630..1205
/gene="rpL1"
/codon_start=1
/translation_table=11
/product="ribosomal protein L1"
/protein_id="AAK17072.1"
/db_xref="GI:13272289"
/translation="MKSIENPLNLTNKKKKNFISIDNLILNKKKKYINLSTNLF
YSTNEMKILFISNIQIENNVYIGNPFIDFLKRVNFSKYVDSENLILVKEKIGK
INSLIIVNCKAYKEIEKKNFYKNIINTLVNNNNILNKIASTVYNSMIIKNEFLLI
FNLKNFEFPNNIYFIEKIIYSSTMSKSFLLK"
1202..1642
/gene="rpL10"
1202..1642
/gene="rpL10"
/codon_start=1
/translation_table=11
/product="ribosomal protein L10"
/protein_id="AAK17073.1"
/db_xref="GI:13272290"
/translation="MMSYNIITKNIKNPLNKVYIILFYNEKIKFELIDNKKKNYK
KEFPCSNKITEFIVPKYSKSLILKHNKNDVYSLINNINDLIDKFLILINKYQV
YDLIENSKLSKENLILKLNKLTILKLNILKNKQYK"
1620..1961
/gene="rpL7/L12"
1620..1961
/gene="rpL7/L12"
/codon_start=1
/translation_table=11
/product="ribosomal protein L7/L12"
/protein_id="AAK17074.1"
/db_xref="GI:13272291"
/translation="MENNIVNVLDISIKNLVEITTLISLEKKNLTVSNEKIAEK
NKFDTLKEIGNKMLIKITIDTITNLNKEKKLVNKNLSIKKNVKKKDEYK
ITLGGIVEIK"
1958..5761
/gene="rpOB"
1958..5761
/gene="rpOB"
/codon_start=1
/translation_table=11
/product="beta subunit of RNA polymerase"
/protein_id="AAK17075.1"
/db_xref="GI:13272292"
/translation="MNLKFNKFCFSKKRPHKCNLPYLLHNOIGSYNSFLSKYNS
LFSIKVLYKOYPLICNNKNIFIKLKLILPCNTEKTKRNHLFSTLYMSIY
VIRKLNLYKIPILNIPSMKKNFIINGIDRILISQFSGYIEYEEKKKKCI
IPLKSGMLEIITNNFLIYEDKINEIWFILGYNKKYFENFEFKIKIRIKG
KRLEPFLNKKKYKLYNKYICISIKILGKLNKFLKNNPLKNNPLKNNIINVNL
LNPIYFPTIPLKKTINLYILYIKIKLKNKILKINKELOQKGYIINFRNLFYKKY
FYSLIAGKRFKRLPKNNKLTIIYEITIKLKTIKENIDONDNLNKLILKCGK
LSTKEDFELKIKYIKINKNNEKKYKDLDIIVSDIITIGLKDYENNELSQFLDQ
NNPLAISHNRKISLISGIEKENGPIRIDIHSHCYKCPIDTPGHNIGLNSL
AYLSKYNKVFISTIKYQFLGILGITFLDKPFRKFIYFNSTIETIYIGIERSP
YFPAKNTNYYKKPLNIDILIEIGDOIISVAGSLIPLSHNDARCLGMSMOQAV
PLIDSENPVIGTGNELIGLNSNYNILDNLNGVLYSDNYKTIKNNNFITVPEKY
TRNNOMTILNOYKVLKGDGVYGNILADSNSTNGEISLCKNLVAPMSYGVFED
SILLSSSILKNNFNSIHIEFTYLVAKIENGFIVSNECFRSNKKINKNVNGIKI
GEFVFSKDVYGMIPKKRKFSPKELKIVFSSSNFYEQPLTVBNKIGITIAV
NDRKIFYPKRIEKLLEPOLNVTCKNNINFEYFNLYIKIKLIPNNKIGITIAV
INSYINININIEKIKCNKKNIFKNIFKNISNELKKNIFYKKNIFKIHDFEN
SIRIKIRIAVKKOJAIGDKSGGNGKGVNSIIDENMYDKGKNKIDLPGLG
VPSRMVGOLEVEFLASNLISKFELKIKNKTISYFKNKLIPIKCIYKNNIKN
NIFNNISLVIKIEKNIKNOLNVCVHNPNYNNKVNIIKTIINKNCLLIPDGTGK
RITQLVNAGITIPKLNHLVIDKTIYSRISGPISTYQOPLGSKSNLUGORGENEYVA
LEAVGAFFLKEMLTITKSDIIGRIELVNNIIRKINDANSIGPESFOYLMKEVSLCP
DIRIL"
5758..9636
gene
/translation="MKILKFKLILKPKAPSSGIGPILGOYGINLIDECNKNFNL
TKFICLEKINVKIAIFDNDNSYOILINTMSLSYIEKKNILISFSKPKPGYNQDTNINLN
DVEIISYFERNIEKKKLNLSRKMMLGTLISMGYENKKYK"
CDS
630..1205
/gene="rpOB"
630..1205
/gene="rpOB"
/codon_start=1
/translation_table=11
/product="beta-prime subunit of RNA polymerase"
/protein_id="AAK17076.1"
/db_xref="GI:13272293"
/translation="MKSIFFKVSIRKMSASSKILISMSFGVTHSNFINKELKPEPN
LFCIKTISDFELCCKKQCFCKDEKLNKRAHNGYKIKIIPKNIPIHTIYSSSS
NISITNLNSKRIELKIKNKLKYVIRKSDIKRKNELIPEKRIPIHTIYSSGAKK
KLSDRELITDCLIKIRIKCNNSFKLPIKIKNLILYIGKNNPSMWAKTQV
LPPRMPPLALSEKRRASSDINELKIVENRYKRRKNTYLDKQOLINRISLOIS
IYALFENKILINPILFSSKRVINELKYSILKGYFRRKNTYLDKQOLINRISLOIS
LFLYECKIYIYALDELKILYELKRRKIIITTSIFIDYDKKNNKQIILKKNVS
YFIMLRAPDLTIRNMFOSPKILLTLEDKTIKIPPLICSLYNDFODQDQAHILPLTNN
KVSNNYLLSINNIIISPNSGNAIITPQDIIINGIYFEDFKNSISLNDIINIF
EPNNVPOINIKENNNKINTPIIGRIINYLPEKKEDYFLMWYKRRKILYIYKILSD
FNLSKIIILKELKIGLITSTYGITIISFDLIEVKNLILNKVFKLYNNIIS
IVELNNPFVKIITIKIQINKTGLKRTNNLFLMDSGSGSMLOIKQILARGFSSK
NGDILNDPIEDNLKNGLNKNKNYEISFGARKGLDTSUKTANSGLYRKLIDVSHDY
IYKIDCKRTGIKIFLNSYFTLVELYKMLYGLILSYLKNKILKENTLIDKMYLF
LILKIKINTIEVRSYFECISNRGACGCGYIDLSNKLIVLIGSVGIIISAGISGPG
OLMFRTEHTGVAAYEFYEDNLIIISNGFVRKKCKVCYNRGEVLIYSLTGEMIRN
KILKELKRYGKIKIFRNGELIKKNTKILINNDNENFIVSNDGYVFKKENNVKN
YCLDNKPIYKTLKNTKITIKSKKNYFIPKNNIYVYCDIIPQELIYAKVSY
ILKSSITGLPRLSELFEARIPDLALISEIDGVCKIKFSHLNITITISFEGREY
TISCLRKLYINNGDYKIDILSDGRPLDNLINDISIDLSYFVEINSIYEPONI
YVNCXKILIELIKOMTKKVIIFISGSSCOQDILFLEDAIENENSTIYNSRLSYVR
IYGTIRITLESLSFESASFOETKILVNSAKIKRDTDYLLGKENVLYGRILPAGTG
LIKYPFKKSSKNNLDFKRRKKFYINVF"
9638..10000
/gene="rps12"
9638..10000
/gene="rps12"
/codon_start=1
/translation_table=11
/product="ribosomal protein S12"
/protein_id="AAK17077.1"
/db_xref="GI:13272294"
/translation="MTNLQILKFKRRKSVKRRKTPALLISSPOKKGICIKVYTTPKP
NSALRRVCVRLSKNNKEITAYIPGEGHNLQESHNVLYGVRKDLPGVKYHILNVYD
LSCGVINRKTSSRSKYGKK"
10000..10470
/gene="rps7"
10000..10470
/gene="rps7"
/codon_start=1
/translation_table=11
/product="ribosomal protein S7"
/protein_id="AAK17078.1"
/db_xref="GI:13272295"
/translation="MSRKRRKRYIILNDPKRGVYIIAKFISYIMNNCKKNLAQKIEY
VSLISITILKNPLIKILIKIYNQGPDEIKKKKIGGSYKIPIKIKLKSLSWPSK
WYKNSKLNENSGYKKNLGVGLIDISTYNNLSSTOKDELNNIIDONKAYSNFKI"
10467..12512
/gene="fusa"
10467..12512
/gene="fusa"
/codon_start=1
/translation_table=11
/product="elongation factor G"
/protein_id="AAK17079.1"
/db_xref="GI:13272296"
/translation="MNDIKININIGIIAHVDAGKTTTERILFSGFSHKIGEVHTG
NITIDMKQOEBGRTITSASVYFKNPNYNSLIDTPGVHPTTIEVRSVLVQ
AVILICASSGIDPQETVNOSEKNNIKILPVNLDLIGAKYIILENIKRRFKN
LIININIGIENSRSQIIDIINKEILNNSOILFENITNNKPDOSNKRNTLSTLS
YDPIIEKINNSFKDITIESIRKVLVINKITIPAGCSSLKNKGIEFLDSIVNLF
SPDIDIKVNSNINSVNSKSFLLERKVDPLGLKIRLERKSEIYSGIEGQIIFNN
SKNIKEKIRIRIIRMANSKRDNLINASAGIYVLIOLKNSFTGDTLSPDNEKVLKELIN
IPLPVIISVEPIYKNDYERKLNILNKCKEEDPSLFFKINENGTGLISGELGELI
IIDRINNEFNIKITKSPQVYSKESIKTKIIGEKYIKQTGRCQYGVYVLKIPILI

```

gene

DR P-PSDB: AAG68236.

XX New polypeptide for creating plant resistance to injury, comprises a
PT gene, the expression of which is induced by stress
XX
XX
PS Claim 4; Page 8; 11pp; Japanese.

CC The present sequence represents a gene isolated from tobacco, in which
CC the expression is induced by stress, and so encodes a stress induced
CC protein (1). The present invention also describes: (1) providing
CC resistance against stress by introducing the above gene to a plant; and
CC (2) a transformed plant which obtained resistance against stress by
CC introducing the above gene to the plant. The gene can be used for
CC creating plants resistant against injury, osmotic pressure, salt, or
CC low temperature-induced stress.

XX Sequence 1210 BP; 347 A; 217 C; 251 G; 395 T; 0 other;

Query Match 100.0%; Score 1210; DB 24; Length 1210;
Best Local Similarity 100.0%; Pred. No. 1.3e-268;
Matches 1210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 tatatcaattgaaacatgctgacaaagagcgcgttccgttggttggttctg 60
DB 1 tatatcaattgaaacatgctgacaaagagcgcgttccgttggttggttctg 60
OY 61 tgaactataagcagttcctaagcagagatattctcaatgtgtcccttctcgcg 120
DB 61 tgaactataagcagttcctaagcagagatattctcaatgtgtcccttctcgcg 120
OY 121 gtgatattcaataaatttccttcctcgtatgagactgtcccgagcattggtgagac 180
DB 121 gtgatattcaataaatttccttcctcgtatgagactgtcccgagcattggtgagac 180
OY 181 gggatattgagctgctgacgtgacagacacaaacggtgtcaattcaatccagaatt 240
DB 181 gggatattgagctgctgacgtgacagacacaaacggtgtcaattcaatccagaatt 240
OY 241 tcgacgtacaggaattacacagagacgtacacaaacggtgtcaattcaatccagaatt 300
DB 241 tcgacgtacaggaattacacagagacgtacacaaacggtgtcaattcaatccagaatt 300
OY 301 atgacagagagaaatttcgaagttttccaaatccaaagcgaatttgatgacatga 360
DB 301 atgacagagagaaatttcgaagttttccaaatccaaagcgaatttgatgacatga 360
OY 361 ctgacgaatcttggatgggttcgtgttaacaatgatatacactatgcaactgtctag 420
DB 361 ctgacgaatcttggatgggttcgtgttaacaatgatatacactatgcaactgtctag 420
OY 421 ctctataaatctgcacagatatactcctaagttttgtagcaaaattcaacgggt 480
DB 421 ctctataaatctgcacagatatactcctaagttttgtagcaaaattcaacgggt 480
OY 481 ttagctacactgtcacaagaagaatattgcagacgttcggatgtgctgcgctgtag 540
DB 481 ttagctacactgtcacaagaagaatattgcagacgttcggatgtgctgcgctgtag 540
OY 541 ttgaactgtgtcagtgctcctgcctccaggaatttcacaaacgaagtcgttactat 600
DB 541 ttgaactgtgtcagtgctcctgcctccaggaatttcacaaacgaagtcgttactat 600
OY 601 caagcacacatcaagcctgtgctatgtgttgagcttcttcgaaacgtaactcttat 660
DB 601 caagcacacatcaagcctgtgctatgtgttgagcttcttcgaaacgtaactcttat 660
OY 661 gtgaaattcgacacgaagtcgtgggggtgaatgacatttgaagaaacacacgacag 720
DB 661 gtgaaattcgacacgaagtcgtgggggtgaatgacatttgaagaaacacacgacag 720
OY 721 ctactgtcgttattgtgcaagagacattcagtttcgaagccttaagcttcagatga 780
DB 721 ctactgtcgttattgtgcaagagacattcagtttcgaagccttaagcttcagatga 780

```

```

DB 721 ctactgtcgttattgtgcaagagacattcagtttcgaagccttaagcttcagatga 780
OY 781 aagtcgactactctcgttattgtatgttcttcttgccggtatagaatagtgagtt 840
DB 781 aagtcgactactctcgttattgtatgttcttcttgccggtatagaatagtgagtt 840
OY 841 tggcgtcaagattctactaggaattccaaatttgatcgaagagtggtgagtgca 900
DB 841 tggcgtcaagattctactaggaattccaaatttgatcgaagagtggtgagtgca 900
OY 901 aaagacgaatttgcatacactcctccgatgaaacagacgtttaagaattttgtagtcaa 960
DB 901 aaagacgaatttgcatacactcctccgatgaaacagacgtttaagaattttgtagtcaa 960
OY 961 gcttttcaagaagtttgcgtatatttttcgaagaactggcaagtgtaacctgtg 1020
DB 961 gcttttcaagaagtttgcgtatatttttcgaagaactggcaagtgtaacctgtg 1020
OY 1021 ctggcgatttaattactaggaattttcttccagtttagtcctgtatttatttgaattc 1080
DB 1021 ctggcgatttaattactaggaattttcttccagtttagtcctgtatttatttgaattc 1080
OY 1081 ttacctattgattgtgatgatttttttcttaaaatttataatttcccaattcttg 1140
DB 1081 ttacctattgattgtgatgatttttttcttaaaatttataatttcccaattcttg 1140
OY 1141 taagtaattgaaatgatatctgacttctcgtcaataatagacaagaacattcgcaaaa 1200
DB 1141 taagtaattgaaatgatatctgacttctcgtcaataatagacaagaacattcgcaaaa 1200
OY 1201 aaaaaaaaaa 1210
DB 1201 aaaaaaaaaa 1210

```

RESULT 2
ABAO1664
ID ABAO1664 standard; DNM: 927 BP.
XX ABAO1664;
XX 06-FEB-2002 (first entry)
DE Nicotiana tabacum cv. Xanthine stress induced gene SEQ ID NO:2.
XX Nicotiana tabacum cv. Xanthine; stress induced protein; plant;
KW stress induced gene; transgenic plant; resistance; ds.
XX Nicotiana tabacum.
OS
FH Key
FT 1.927
FT CDS
FT Location/Qualifiers
FT /tag= a
FT /product= "stress induced protein"
FT /trans_except= (pos:828..830,aa:Xaa)
PN JP2001252084-A.
PD 18-SEP-2001.
PF 15-MAR-2000; 2000JP-0071655.
PR 15-MAR-2000; 2000JP-0071655.
PA (NARA-) NARA SENTAN KAGAKU GIJUTSU DAIGAKUIN DAI.
DR WPI: 2002-029660/04.
XX P-PSDB: AAG68236.
PT New polypeptide for creating plant resistance to injury, comprises a
XX gene, the expression of which is induced by stress
XX
XX Claim 3; Page 7-8; 11pp; Japanese.

XX The present sequence represents a gene isolated from tobacco, in which
 CC the expression is induced by stress, and so encodes a stress induced
 CC protein (1). The present invention also describes: (1) providing
 CC resistance against stress by introducing the above gene to a plant; and
 CC (2) a transformed plant which obtained resistance against stress by
 CC introducing the above gene to the plant. The gene can be used for
 CC creating plants resistant against injury, osmotic pressure, salt, or
 CC low temperature-induced stress.

SX Sequence 927 BP; 256 A; 182 C; 212 G; 277 T; 0 other;

Query Match 76.6%; Score 927; DB 24; Length 927;
 Best Local Similarity 100.0%; Pred. No. 1.3e-203;
 Matches 927; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 atgttgacaaagaggcgtcttctgctgtgtgttcttactgtgacacataagcagt 77
 DB 1 atgttgacaaagaggcgtcttctgctgtgtgttcttactgtgacacataagcagt 60
 QY 78 tctaaagcagagatatttccatgtgtcccttcttccgtgagtgatattcaataaa 137
 DB 61 tctaaagcagagatatttccatgtgtcccttcttccgtgagtgatattcaataaa 120
 QY 138 ttcccttcgcagctgagagctgataccgagcattgtgtaagcagcagataagctcgat 197
 DB 121 ttcccttcgcagctgagagcctgataccgagcattgtgtaagcagcagataagctcgat 180
 QY 198 tgcacgaagaacaaacccgtgttcaataacacacgaattttcgcagcgaagaaatt 257
 DB 181 tgcacgaagaacaaacccgtgttcaataacacacgaattttcgcagcgaagaaatt 240
 QY 258 aactacagaagctactacataaagcctactgtcctgagcctaaatgtatcagaagaaat 317
 DB 241 aactacagaagctactacataaagcctactgtcctgagcctaaatgtatcagaagaaat 300
 QY 318 tgcacagtttttccaatcacagcagcaggttactgtatgcacagctagcacaattttgaa 377
 DB 301 tgcacagtttttccaatcacagcagcaggttactgtatgcacagctagcacaattttgaa 360
 QY 378 tgggttcgtgttaacaatgatacaactgtcaactgtcagcttactcaatctgtca 437
 DB 361 tgggttcgtgttaacaatgatacaactgtcaactgtcagcttactcaatctgtca 420
 QY 438 cagtatatcctacaaagtttttagcacaataatcaacgggttttagctacactgtcata 497
 DB 421 cagtatatcctacaaagtttttagcacaataatcaacgggttttagctacactgtcata 480
 QY 498 agagaataatgcaagcttgagattgctgctgctgctgtgaggttgaactgtgtcattg 557
 DB 481 agagaataatgcaagcttgagattgctgctgctgctgtgaggttgaactgtgtcattg 540
 QY 558 tccctgtctcagagcatttcaacaacagcgtctacagtatataacacacatcagcggc 617
 DB 541 tccctgtctcagagcatttcaacaacagcgtctacagtatataacacacatcagcggc 600
 QY 618 ctggcttatgggtttgagccttcttggagaagcgtatctgtatgtataaatttgcagccg 677
 DB 601 ctggcttatgggtttgagccttcttggagaagcgtatctgtatgtataaatttgcagccg 660
 QY 678 agtcgttgggggtgagtgacatttaagaagaacagcagcagagctatgtcgtatttg 737
 DB 661 agtcgttgggggtgagtgacatttaagaagaacagcagcagagctatgtcgtatttg 720
 QY 738 tgcagaagaggaatccaagtttcgaagcttaagtcgagtgcaagtcgaagctactatct 797
 DB 721 tgcagaagaggaatccaagtttcgaagcttaagtcgagtcgaagtcgaagctactatct 780
 QY 798 gtctatgtatgttcttggcgggtatagaagaagtgaggttgggcgaagaatttcta 857
 DB 781 gtctatgtatgttcttggcgggtatagaagaagtgaggttgggcgaagaatttcta 840

QY 858 ctagaatctccatctctgtatctgcagcagtggtgtgacagtgcaaaagcgaatttgcatt 917
 DB 841 ctagaatctccatctctgtatctgcagcagtggtgtgacagtgcaaaagcgaatttgcatt 900
 QY 918 acatctcccgatgacacgaactgttaa 944
 DB 901 acatctcccgatgacacgaactgttaa 927

RESULT 3

AAH71471/C
 ID AAH71471 standard; cDNA; 612 BP.

AC AAH71471;

DX 19-SEP-2001 (first entry)

DE Human cervical cancer marker nucleic acid 2745.

KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

OS Homo sapiens.

PN W0200142467-A2.

PD 14-JUN-2001.

PF 08-DEC-2000; 2000WO-US33312.

PR 08-DEC-1999; 990US-0169681.

PR 21-DEC-1999; 990US-0171350.

PR 14-MAR-2000; 2000US-0189315.

PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210600.

PR 21-JUL-2000; 2000US-0220114.

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Schlegel R, Deeds J, Berger A, Zhao X;

XX

DR WPI; 2001-375006/39.

XX

PT New isolated nucleic acid for diagnosing and treating cervical cancer
 and for assessing and detecting compounds for treating the cancer -

XX

PS Claim 1; Page 564; 1051pp; English.

XX

CC The invention relates to novel genes (AAH68727-AAH73383) associated with
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded
 CC polypeptides are useful to assess if a patient is afflicted with
 CC cervical cancer or has a pre-malignant condition; to monitor the
 CC progression of cervical cancer or a premalignant condition in a patient;
 CC and to select and/or assess the efficacy of a compound or therapy for
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be
 CC useful for gene therapy.

CC

XX

SQ

Sequence 612 BP; 380 A; 50 C; 2 G; 178 T; 2 other;

Query Match 5.1%; Score 62.2; DB 22; Length 612;
 Best Local Similarity 48.1%; Pred. No. 6.1e-05;
 Matches 204; Conservative 0; Mismatches 219; Indels 1; Gaps 1;

QY 787 agtaactatctggttaatgtatgttcttggcgtatagaagaatgaggttttgccgc 846
 DB 461 AATTTTAAATTTTATTTTATTTTGGTTAGGTTATATGATATGATGAAGA-TTTTGAAA 403
 QY 847 taagattctactaagaatccaatctgtatcgacagcagtggtgtgacagtgcaaaagac 906
 DB 402 TGTATTTTGTGTTTTTTTTTTTATTTTATTTTATTTTATTTTATTTTAAAGAT 343
 QY 907 ggaatttcataacatccctcgatgaacagacgcttaagaatttttctagtaagcattt 966

Db 342 TAAATGTTTTTTTTTTTAAATTTTGTGTTTTTTTTTTTTTTTTTTTATT 283
QY 967 ttaacagaatttctgtatattttcagaanaactagacaagctacgttgcgcg 1026
Db 282 TTAAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 223
QY 1027 attaactacagattttcttcacgttagtcgcgttatttattgatttact 1086
Db 222 GGGTATGGGTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 163
QY 1087 attgattgtaataattttctcctaataatttataatttccattctgaagta 1146
Db 162 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 103
QY 1147 attgaatggaatttcttctctctcgaataatagaacacattcgcaaaaaa 1206
Db 102 ATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 43
QY 1207 aaaa 1210
Db 42 AAAA 39
RESULT 4
AAL20340
ID AAL20340 standard; cDNA; 622 BP.
XX
AC AAL20340;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 12797.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
FN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PE 10-JAN-2001; 2001WO-US00798.
XX
PF 14-JAN-2000; 2000US-0176077.
XX
PR 14-MAR-2000; 2000US-0189167.
XX
PR 24-MAR-2000; 2000US-0192099.
XX
PR 29-MAR-2000; 2000US-0193480.
XX
PR 15-MAY-2000; 2000US-0205230.
XX
PR 09-JUN-2000; 2000US-0211315.
XX
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI: 2001-451856/48.
XX
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 2261-2262; 3693pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterizing treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
CC
SQ Sequence 622 BP; 189 A; 154 C; 91 G; 188 T; 0 other;

Query Match 4.9%; Score 59.6; DB 22; Length 622;
Best Local Similarity 59.4%; Pred. No. 0.0024;
Matches 101; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 1041 tttttcttcacgttagtcgcgttatttattgatttacttaacttattgattg 1100
Db 46 ttt 105
QY 1101 gattttttctcctaataatttataatttccattctgtaagaattgaatgata 1160
Db 106 ttt 165
QY 1161 tgaacttctgtaataatagaacacattcgcaaaaaa 1210
Db 166 tttcttgagggaataaaaaaaaccacaaaaaa 215
RESULT 5
AAL8595/C
ID AAL8595 standard; cDNA; 411 BP.
XX
AC AAL8595;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 8655.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
FN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PE 26-FEB-2001; 2001WO-US04927.
XX
PF 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAR-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-514838/56.
XX
DR P-PSDB: AAC08664.
XX
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
XX
PS Claim 1; SEQ ID NO 8655; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAL79941-AAL93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 411 BP; 214 A; 24 C; 64 G; 107 T; 2 other;


```

PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI: 2001-451856/48.
XX
XX New peptide useful as a marker for the diagnosis of breast cancer -
XX
XX Claim 1; Page 473; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
XX (AA107544-AA126789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterizing treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity.
XX
XX Sequence 394 BP; 218 A; 11 C; 16 G; 93 T; 56 other;
XX
XX
XX Query Match 4.3%; Score 52.6; DB 22; Length 394;
XX Best Local Similarity 49.7%; Pred. No. 0.0086;
XX Matches 91; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
XX
XX 1028 ttaactagatttcttccagtttagctgtatttattgatatcttactaa 1087
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 222 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 163
XX
XX 1088 ttgtgtgtagatgttttttcttcttaaaatttataatttcttcaattctgtaagtaa 1147
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 163 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 103
XX
XX 1148 ttgaatgagatttcttctgtcaataatagaacaagacattcgcaaaaaa 1207
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 102 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 43
XX
XX 1208 aaa 1210
XX |||
XX 42 Aaa 40
XX
XX
XX RESULT 12
XX AAS60450
XX ID AAS60450 standard; CDNA: 425 BP.
XX
XX AAS60450;
XX
XX 29-JAN-2002 (first entry)
XX
XX Human cancer agent-sensitive marker #181.
XX
XX Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
XX squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
XX lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
XX Hodgkin's disease; glioma; ss.
XX
XX Homo sapiens.
XX
XX WO200179556-A2.
XX
XX 25-OCT-2001.
XX
XX 13-APR-2001; 2001WO-US12132.
XX
XX 14-APR-2000; 2000US-197538P.
XX

```

```

XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Brown JL, Bolt A, Van Huffel C;
XX WPI: 2001-602933/68.
XX
XX Novel nucleic acid, used as a marker to determine the effectiveness of
XX using TAXOL to treat cancer cell growth in individuals -
XX
XX Claim 1; Page 221; 527pp; English.
XX
XX The invention relates to 1046 novel nucleic acids which are used as
XX markers for determining the sensitivity of a cancer cell to the
XX anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
XX they are shown to express one of the 242 sensitivity markers or the
XX cells are shown not to express one of the 804 resistance markers.
XX The methods can be used to determine the effectiveness of TAXOL
XX in the treatment of cancer cell growth in an individual. The markers
XX can be used as targets in developing anti-cancer agents such as
XX chemotherapeutic compounds. The markers can also be used as targets in
XX developing treatments for cancer, particularly those cancers which
XX display resistance to agents and exhibit expression of the markers. The
XX anticancer agents developed by the novel method can be used to treat
XX cancer. Probes based on the markers can be used to detect transcripts or
XX genomic sequences corresponding to the markers, in the identification of
XX cells or tissues which mis-express the protein. Cancers which may
XX be targeted include carcinoma (e.g. squamous cell carcinoma),
XX sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic leukaemia),
XX lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and
XX tumours (e.g. glioma). The present sequence is one of the 1046
XX novel cancer cell markers.
XX
XX Sequence 425 BP; 159 A; 41 C; 28 G; 194 T; 3 other;
XX
XX
XX Query Match 4.3%; Score 52.4; DB 22; Length 425;
XX Best Local Similarity 55.4%; Pred. No. 0.0099;
XX Matches 98; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
XX
XX 1034 actagatttcttctcagatttagctgtatttattgatatcttaccatttgat 1093
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 1 acttttttttttttttttttttttttttttttttttttttttttttttttttttt 60
XX
XX 1094 tctgtagatttttcttcttaaaatttataatttcttcaattcttgaagtaattgaat 1153
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 61 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 120
XX
XX 1154 ggaatttgaatttctgtcaataatagaacaagacattcgcaaaaaa 1210
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 121 aaaaattttaaataaaaaaataaaaaaatttcccccaaaaaa 177
XX
XX
XX RESULT 13
XX AAL11441
XX ID AAL11441 standard; CDNA: 426 BP.
XX
XX AAL11441;
XX
XX 07-DEC-2001 (first entry)
XX
XX Human breast cancer expressed polynucleotide 3898.
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
XX
XX Homo sapiens.
XX
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US00798.
XX

```

PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX
XX New peptide useful as a marker for the diagnosis of breast cancer -
XX
XX Claim 1; Page 712; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
CC (AA07544-AA06789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterizing treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
XX Sequence 426 BP; 81 A; 57 C; 36 G; 159 T; 93 other;
SQ
Query Match 4.3%; Score 52.4; DB 22; Length 426;
Best Local Similarity 50.6%; Pred. No. 0.01;
Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 1041 tttttccagtttagccgcgtgatttattgatatcttaacctatttgatggtat 1100
DB 45 ttt 104
QY 1101 gatcttttccctaatttataatttccctaattctgtgaagtgaatgatatt 1160
DB 105 ttt 164
QY 1161 tgcacttctgtcaatacagaacagacattcgcaaaaaa 1210
DB 165 ttttttng99gnaaaaaaannnaaaanaaaaa 214
RESULT 14
ABL32513
ID ABL32513 standard; DNA; 8805 BP.
XX
XX ABL32513;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 486.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antiataemic; cytosclastic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antididiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX MO200200928-A2.
PN
XX
XX 03-JAN-2002.
PD
XX

PE 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
XX
XX Claim 1; SEQ ID NO 486; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX Sequence 8805 BP; 2030 A; 477 C; 2377 G; 3921 T; 0 other;
SQ
Query Match 4.3%; Score 52.4; DB 24; Length 8805;
Best Local Similarity 50.4%; Pred. No. 0.021;
Matches 128; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 939 ttttaagatttttgcagcagcatttttaacagaatttggttttttcagaaa 998
DB 6722 tgcattttaaatttaataaataaataaataaatttttttggttggaattt 6781
QY 999 tctagacaagtcacacccgtgctgcgtgataattactaggatttttccagtttag 1058
DB 6782 ttaagaaagtttttttttttttggagtggtgttttttgattttttttgttattt 6841
QY 1059 tctgtattttattgatatcttacctatttgattgattgatttttccctaaat 1118
DB 6842 tgcgtatta 6901
QY 1119 ttataatttccctaattctgtgaattgaatgaatgatttgaatttgcataaa 1178
DB 6902 gtttttttttttataattgattttttttttttttttttttttttttttttttttt 6961
QY 1179 tagaacaagacatt 1192
DB 6962 ttatttaagtgtt 6975
RESULT 15
AAS61101
ID AAS61101 standard; DNA; 12592 BP.
XX
XX AAS61101;
XX
XX 29-JAN-2002 (first entry)
XX
XX Human gene regulation-associated gene oligonucleotide #56.
DE
XX
XX Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saeche-Chozen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
OS
XX Homo sapiens.
XX

Search completed: June 4, 2002, 21:31:08
 Job time: 3816 sec

```

XX PN NO20017375-A2.
XX XX 18-OCT-2001.
XX XX 06-APR-2001; 2001WO-EP03968.
XX PF 06-APR-2000; 2000DE-1019058.
XX PR 07-APR-2000; 2000DE-1019173.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPiG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-017470/02.
XX DR
XX PT New nucleic acid sequences from chemically modified genes associated
XX PT with gene regulation, useful for analysing cytosine methylations for
XX PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
XX PT disease
XX PS Claim 1; SEQ ID No 57; 26pp; English.
XX PS
XX CC The invention relates to 224 nucleic acid sequences comprising at least
XX CC 18 bases of a chemically pretreated gene associated with gene regulation
XX CC selected from 43 known genes (or complementary sequences). The
XX CC chemical pretreatment converts cytosine bases unmethylated at the
XX CC 5-position to uracil or another base with hybridisation behaviour
XX CC dissimilar to cytosine, to enable analysis of cytosine methylations.
XX CC The DNA sequences, oligomers (or sets/arrays) and method are
XX CC useful in the diagnosis of diseases (or predisposition to diseases)
XX CC associated with gene regulation and in therapy of such diseases, by
XX CC enabling analysis of the cytosine methylation patterns of such genes,
XX CC kits are provided. They are especially useful in diagnosis
XX CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
XX CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
XX CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
XX CC pre-eclampsia, graft versus-host disease. The present sequence is a
XX CC sequence included in the sequence data for this specification and is
XX CC associated with the human gene regulation-associated genes.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 12592 BP: 3783 A; 195 C; 2736 G; 5878 T; 0 other;
XX
Query Match 4.3%; Score 52.2; DB 24; Length 12592;
Best Local Similarity 53.9%; Pred.No.0.026;
Matches 130; Conservative 0; Mismatches 108; Indels 3; Gaps 1;
QY 939 tgttaagattttgtcagtaagcattttaacaagaattgtcgtatttttcagaanaa 998
DB 6055 tatttgaagttaagaataattgtatttttagaattttttatatattgttttaaaag 6114
QY 999 tctagagcaagcgcgtgctggcgatttaatactaggaattttcttcaggttag 1058
DB 6115 tttagagta--ttaaatagaataatttaataataattattttatattattat 6171
QY 1059 tccgtattttattatattcttaacctatttgattgtgtacgatttttcccttaaat 1118
DB 6172 tttttaaagataagaataattatttttttagttgtgttttttttagttggtta 6231
QY 1119 ttataaatttcctaattctgtgaagtaattgaatgataattgtacttctgtcaataa 1178
DB 6232 tataatagattattagttgtttgaataaacaagtaagattatatttgaattta 6291
QY 1179 t 1179
DB 6292 t 6292

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: June 4, 2002, 20:26:47 ; Search time 49.17 Seconds

(without alignments)
604.872 Million cell updates/sec

Title: US-09-805-919-3

Perfect score: 1210

Sequence: 1 tatattcatgtgaacatg.....ttcgcaaaaaaaaaaaaaa 1210

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents, NA.*
1: /cgn2_6/p/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/p/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/p/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/p/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/p/ptodata/2/ina/PCRTUS_COMB.seq.*
6: /cgn2_6/p/ptodata/2/ina/backfillseq1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1 | 49.4 | 4.1 | 9636 | 1 US-08-323-170B-1 | Sequence 1, Appl |
| 2 | 49.4 | 4.1 | 9636 | 4 US-08-954-441-1 | Sequence 1, Appl |
| 3 | 45.8 | 3.8 | 6243 | 3 US-09-056-075-1 | Sequence 1, Appl |
| 4 | 44.6 | 3.7 | 152331 | 3 US-09-128-155-16 | Sequence 16, Appl |
| 5 | 44.4 | 3.7 | 6243 | 2 US-09-056-075-1 | Sequence 1, Appl |
| 6 | 44.4 | 3.7 | 19124 | 2 US-08-487-826B-13 | Sequence 13, Appl |
| 7 | 44.2 | 3.7 | 3926 | 2 US-08-731-722-1 | Sequence 1, Appl |
| 8 | 44.2 | 3.7 | 3926 | 2 US-08-731-722-2 | Sequence 2, Appl |
| 9 | 44.2 | 3.7 | 3926 | 2 US-08-731-722-3 | Sequence 3, Appl |
| 10 | 43.8 | 3.6 | 2447 | 2 US-09-014-969-14 | Sequence 14, Appl |
| 11 | 43.6 | 3.6 | 1066 | 1 US-08-157-101A-4 | Sequence 4, Appl |
| 12 | 43.6 | 3.6 | 1882 | 4 US-09-370-253-1 | Sequence 1, Appl |
| 13 | 43.2 | 3.6 | 1117 | 4 US-08-247-373B-33 | Sequence 33, Appl |
| 14 | 43 | 3.6 | 176373 | 3 US-09-128-155-17 | Sequence 17, Appl |
| 15 | 42.8 | 3.5 | 5852 | 1 US-07-867-106-2 | Sequence 2, Appl |
| 16 | 42.6 | 3.5 | 2852 | 3 US-09-027-137-2 | Sequence 2, Appl |
| 17 | 42.4 | 3.5 | 2674 | 4 US-09-817-180-1 | Sequence 1, Appl |
| 18 | 42 | 3.5 | 240 | 1 US-08-628-417-6 | Sequence 6, Appl |
| 19 | 42 | 3.5 | 1454 | 4 US-09-372-422A-19 | Sequence 19, Appl |
| 20 | 41.8 | 3.5 | 658 | 4 US-08-998-416-595 | Sequence 595, Appl |
| 21 | 41.6 | 3.4 | 1172 | 1 US-07-945-288-9 | Sequence 9, Appl |
| 22 | 41.6 | 3.4 | 1172 | 1 US-08-462-831-9 | Sequence 9, Appl |
| 23 | 41.6 | 3.4 | 1172 | 1 US-08-461-809-9 | Sequence 9, Appl |
| 24 | 41.6 | 3.4 | 1172 | 1 US-08-461-441-9 | Sequence 9, Appl |
| 25 | 41.6 | 3.4 | 1172 | 5 PCT-US93-08518-9 | Sequence 9, Appl |
| 26 | 41.6 | 3.4 | 1474 | 4 US-08-821-994-64 | Sequence 64, Appl |
| 27 | 41.6 | 3.4 | 1700 | 2 US-08-897-340-4 | Sequence 4, Appl |

| | | | | | |
|------|------|-----|------|---------------------|-------------------|
| C 28 | 41.6 | 3.4 | 1700 | 3 US-09-252-329-4 | Sequence 4, Appl |
| C 29 | 41.4 | 3.4 | 1582 | 3 US-08-545-196B-10 | Sequence 10, Appl |
| C 30 | 41.4 | 3.4 | 1582 | 3 US-08-545-196B-12 | Sequence 12, Appl |
| C 31 | 40.8 | 3.4 | 1134 | 3 US-09-248-335-29 | Sequence 29, Appl |
| C 32 | 40.8 | 3.4 | 1411 | 4 US-08-964-127-5 | Sequence 5, Appl |
| C 33 | 40.8 | 3.4 | 1411 | 4 US-09-496-692-5 | Sequence 5, Appl |
| C 34 | 40.8 | 3.4 | 5852 | 1 US-07-867-106-2 | Sequence 2, Appl |
| C 35 | 40.6 | 3.4 | 140 | 1 US-08-628-417-5 | Sequence 5, Appl |
| C 36 | 40.4 | 3.3 | 1493 | 1 US-08-340-820-24 | Sequence 24, Appl |
| C 37 | 40.4 | 3.3 | 1493 | 1 US-08-593-535-24 | Sequence 24, Appl |
| C 38 | 40.4 | 3.3 | 3138 | 1 US-07-867-106-4 | Sequence 4, Appl |
| C 39 | 40.2 | 3.3 | 270 | 2 US-08-520-678A-30 | Sequence 30, Appl |
| C 40 | 40.2 | 3.3 | 270 | 2 US-08-897-126-30 | Sequence 30, Appl |
| C 41 | 40.2 | 3.3 | 3763 | 1 US-07-792-865D-1 | Sequence 1, Appl |
| C 42 | 40 | 3.3 | 1349 | 2 US-08-676-783-10 | Sequence 10, Appl |
| C 43 | 40 | 3.3 | 8920 | 2 US-08-446-855A-1 | Sequence 1, Appl |
| C 44 | 40 | 3.3 | 8920 | 2 US-09-150-741-1 | Sequence 1, Appl |
| C 45 | 39.6 | 3.3 | 665 | 2 US-08-883-795A-36 | Sequence 36, Appl |

ALIGNMENTS

RESULT 1
US-08-323-170B-1
Sequence 1, Application US/08323170B
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1
Query Match 4.1% Score 49.4; DB 1; Length 9636;

Best Local Similarity 56.4%; Pred. No. 0.0012;
Matches 92; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 1042 ttctctccagttgagctgctgatttattgataattcttaccatttgattgctatg 1101
Db 12 TTTTATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 71

Qy 1102 atttttccctaaatttatttatttccctaatctctgtaagttaattgaaatgataatt 1161
Db 72 TTTTATTTTATTTATTTTATTTTATTTTCTTCACATCTCTTATCTCTTATCTGATATTT 131

Qy 1162 gtacttctgcaataatagacaagacattcgcaaaaaaa 1204
Db 132 TTTCTTTCTTTTATAATGAGAAATTAACGCTGAGAA 174

RESULT 2

US-08-954-441-1
; Sequence 1, Application US/08954441
; Patent No. 6316000

GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf230
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 20-OCT-1997
APPLICATION NUMBER: US/08/954,441
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
FILING DATE: 13-OCT-1994
APPLICATION NUMBER: US 08/323,170
PRIOR APPLICATION DATA: US 08/010,409

ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 015280-113110US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-954-441-1

Query Match 4.1%; Score 49.4; DB 4; Length 9636;
Best Local Similarity 56.4%; Pred. No. 0.0012;
Matches 92; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 1042 ttctctccagttgagctgctgatttattgataattcttaccatttgattgctatg 1101

Db 12 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 71

Qy 1102 atttttccctaaatttataatttccctaatctctgtaagttaattgaaatgataatt 1161
Db 72 TTTTATTTTATTTATTTTATTTTCTTCACATCTCTTATCTCTTATCTGATATTT 131

Qy 1162 gtacttctgcaataatagacaagacattcgcaaaaaaa 1204
Db 132 TTTCTTTCTTTTATAATGAGAAATTAACGCTGAGAA 174

RESULT 3

US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368

GENERAL INFORMATION:
APPLICANT: Johnson, Eric A.
APPLICANT: Bradshaw, Marile
APPLICANT: Rood, Julian
TITLE OF INVENTION: Expression system for Clostridium
TITLE OF INVENTION: Species
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 09/05/96, 075
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95238
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: misc-feature
LOCATION: 3770..4013
OTHER INFORMATION: /note="RP4 origin of DNA transfer (orig) from
US-09-056-075-1
OTHER INFORMATION: plasmid RP4"

Query Match 3.8%; Score 45.8; DB 2; Length 6243;
Best Local Similarity 54.6%; Pred. No. 0.0093;
Matches 136; Conservative 0; Mismatches 107; Indels 6; Gaps 2;

Qy 964 atttaccagagttggtgattttttttagaaatcctaggaaggttaacctgtgctg 1023
Db 3200 ATATATCTTTATTTATTTATTTGTTTAAAAAAGTGAATTCGGCTTTATTT- 3258

Qy 1024 gcgatttaactagagattttcttccagttgagc---ctgatttatttgataatt 1079
Db 3259 -AAACTATATTTTAGCAATTTTATTTTCATTTTCATCTTACAGATTTGATATATCTT 3317

| | | | |
|----|------|---|------|
| Qy | 1080 | cttcaccatttggtctgctgctggttttttccctaaatcttcaattccctactctt | 1159 |
| | | | |
| Db | 3318 | TAAATAGCTTTATCAAAATATATCTTTTCTCAAAATTTATATATTTTATATATTTA | 3377 |
| | | | |
| Qy | 1140 | gttagctaatgtagatctctgtactctctgcataatatagaacagacattcgcaaa | 1199 |
| | | | |
| Db | 3378 | TTATATATATATTTTATTTTATAGTTCTTCTCAACACGCTTTAAAAAGAACTTTAAA | 3437 |
| | | | |
| Qy | 1200 | aaaaaaaaa | 1208 |
| | | | |
| Db | 3438 | ATTAATAACA | 3446 |

```

RESULT      4
US-09-128-155-16/c
: Sequence 16, Application US/09128155
: Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
: TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
EARLIER FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ. ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ. ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

| | | | | |
|-----------------------|-----------------|--|-----------|----------------|
| Query Match: | 3.7% | Score 44.6: | Db 3: | Length 122331; |
| Best Local Similarity | 48.6%; | Pred. No. 0.069: | | |
| Matches 122; | Conservative 0; | Mismatches 129; | Indels 0; | Gaps 0; |
| QY | 903 | agacgggaatttgcatacatccctccgagtaacagaactgttaagaattttcgtatgaagc | 962 | |
| Db | 23610 | AGATAGGCTTAACAATAACACATCTCTGTATATATAATCTGTAGATAGACATTAATCTTC | 23551 | |
| QY | 963 | tatttcaacagaagtttgtatatttttttccagaacatctagacaaggtcaacctgtgc | 1022 | |
| Db | 23550 | CATTCACATTAATTCGTTTATCTGTGAAAAAGATATGAACTGTCCATCTCTGTGTGAAGAAT | 23491 | |
| QY | 1023 | ggcgattaatctactaagaattttcttccagtttagtcctgtatatttatttgatattcct | 1082 | |
| Db | 23490 | AGAGATTATTTTATAGAGTTTGTCTGTGTGTGGGATCTGTCTTTCTTTTAAATTTG | 23431 | |
| QY | 1083 | acctatttgattgtgtatgatatttttccctaaatattatcaatttcccaattcctgta | 1142 | |
| Db | 23430 | AAATATGTACTCTGTGGAATGATTTTATAAATGATTTTACACATTTTGGAAAGGATATT | 23371 | |
| QY | 1143 | agtaattgaatt 1153 | | |
| Db | 23370 | AATGATAGAAAT 23360 | | |

RESULT 5
US-09-056-075-1/c
: Sequence 1, Application US/09056075
: Patent No. 5955368
: GENERAL INFORMATION:
: APPLICANT: Johnson, Eric A.
: APPLICANT: Bradshaw, Marite

```

1  APPLICANT:  Rood, Julian
2  TITLE OF INVENTION:  Expression System for Clostridium
3  TITLE OF INVENTION:  Species
4  NUMBER OF SEQUENCES:  2
5  CORRESPONDENCE ADDRESSES:
6  ADDRESSEE:  Quarles & Brady
7  STREET:  1 South Pinckney Street
8  City:  Madison
9  STATE:  WI
10 COUNTRY:  US
11 ZIP:  53701-2113
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  Floppy disk
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  PC-DOS/MS-DOS
17 SOFTWARE:  patentin Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/09/056, 075
20 FILING DATE:
21 CLASSIFICATION:
22 ATTORNEY/AGENT INFORMATION:
23 NAME:  Seay, Nicholas J.
24 REGISTRATION NUMBER:  27386
25 REFERENCE/DOCKET NUMBER:  960296.95238
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE:  608-251-5000
28 TELEFAX:  608-251-9166
29 INFORMATION FOR SEQ ID NO:  1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH:  6243 base pairs
32 TYPE:  nucleic acid
33 STRANDEDNESS:  double
34 TOPOLOGY:  linear
35 MOLECULE TYPE:  DNA (genomic)
36 FEATURE:
37 NAME/KEY:  misc.feature
38 LOCATION:  3770..4013
39 OTHER INFORMATION:  /note= "RPA origin of DNA transfer (orit) from
40                    "
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959

```

[illegible]

RESULT 6
US-08-487-826B-13/C
Sequence 13, Application US/08487825B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chintan, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhuan

APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMIDIUM VIAX
TITLE OF INVENTION: AND PLASMIDIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 3.7%; Score 44.4; DB 2; Length 19124;
Best Local Similarity 48.4%; Pred. No. 0.034;
Matches 123; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 929 tgaacagacgtcgaagcttctgctgacgacgtatcttaacagaagctgctgcttctt 988
DB 15977 TGCAGTGAATTAATTTTATTTTATTTTATTAATTAATTAATTTTATTTTATTTTATTTT 15918
QY 988 tticagaaatctagacaagctgacacgtgctgacgtatcttaacagaagcttcttctt 1048
DB 15917 TTTCAATTAATTTTATTTTATTTTATTTTATTAATTAATTTTATTTTATTTTATTTT 15858
QY 1049 tccagcttaagctcgtatcttaattgatacttcaactatctgacgtgctgatacttctt 1108
DB 15857 TTTATTTAATTAATTTTATTTTATTTTATTTTATTTTATTAATTAATTTTATTTTATTT 15798
QY 1109 tccctaaatttataatttccctcaattcctgtaagtaattgagatgatacttcttctt 1168
DB 15797 TTAATGTAATATTTTATTTTATTAACATTTTATTTTATTTTATTTTATTTTATTAATTAATTT 15738
QY 1169 ctgtcaataataga 1182
DB 15737 TTTATTTAATATA 15724

RESULT 7
US-08-731-722-1
Sequence 1, Application US/08731722
Patent No. 5961971
GENERAL INFORMATION:
APPLICANT: Martin, Frank N.
TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens
TITLE OF INVENTION: by Pythium oligandrum
NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/731,722
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF-161
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-3600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3926 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 1982-24
US-08-731-722-1

Query Match 3.7%; Score 44.2; DB 2; Length 3926;
Best Local Similarity 51.2%; Pred. No. 0.021;
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 974 aagttggtatctttcttcgaatctagacaagctgacacgtgctgacgtatctt 1033
DB 1231 AATGTGAGTGATTAATACAGATGCGTTTATTTGAACTGATGAGATTATTTT 1310
QY 1034 actagagatcttcttcgaagttagctcgtatcttaattgatacttcaactatcttgaat 1093
DB 1311 CTTCGAATTTTCTTAATTAATTTTATTCCTATTTTGTATTCGTTTGTCTACTTATATA 1370
QY 1094 tgtatgatttttcttcttaaatcttaatttcaatttccctaatcttgaatgaatgaat 1153
DB 1371 TTAATTAATTTTATTTTCTTAATTAATTTTCAATTTTGAATTTTGTATTTTGTATTTTCTT 1430
QY 1154 ggaatacttgactcttctcgtca 1174
DB 1431 AAATTTCTAAGCTCTTAGCA 1451

RESULT 8
US-08-731-722-2
Sequence 2, Application US/08731722
Patent No. 5961971
GENERAL INFORMATION:
APPLICANT: Martin, Frank N.
TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens
TITLE OF INVENTION: by Pythium oligandrum
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/731.722
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF-161
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3926 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: 1985-5
US-08-731-722-2

Query Match 3.7%; Score 44.2; DB 2; Length 3926;
Best Local Similarity 51.2%; Pred. No. 0.021;
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 974 aagtttgatatttttccagaacactagacaagtcacccgtgctgcgataatt 1033
DB 1251 AATTGGAGTGGATTAGTACGACATCGGTTTATTGGAACCTATCGAGTTTATT 1310
QY 1034 actagagattttctccagtttagtcctgataatttattgatacttaccattgat 1093
DB 1311 CTTGGAATTTTCTAATATTATTCCTCATTTTGTATTCGTTTGTCTACTTATATA 1370
QY 1094 tgtatagatttttccctaaatttaatttcccaattcttgaagtaattgaat 1153
DB 1371 TTAATAAATTTTTCCTAATATTTCATAATTTGATTTGATTTGATATATTCTT 1430
QY 1154 ggaattgtacttctgcga 1174
DB 1431 AAATTTCTACTCTTAGCA 1451

RESULT 9
US-08-731-722-3
; Sequence 3, Application US/08731722
; Patent No. 5961971
; GENERAL INFORMATION:
; APPLICANT: Martin, Frank N.
; TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731.722
; FILING DATE:
; CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF-161
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3933 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 1986-41
US-08-731-722-3

Query Match 3.7%; Score 44.2; DB 2; Length 3933;
Best Local Similarity 51.2%; Pred. No. 0.021;
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 974 aagtttgatatttttccagaacactagacaagtcacccgtgctgcgataatt 1033
DB 1255 AATTGGAGTGGATTAGTACGACATCGGTTTATTGGAACCTATCGAGTTTATT 1314
QY 1034 actagagattttctccagtttagtcctgataatttattgatacttaccattgat 1093
DB 1315 CTTGGAATTTTCTAATATTATTCCTCATTTTGTATTCGTTTGTCTACTTATATA 1374
QY 1094 tgtatagatttttccctaaatttaatttcccaattcttgaagtaattgaat 1153
DB 1375 TTAATAAATTTTTCCTAATATTTCATAATTTGATTTGATTTGATATATTCTT 1434
QY 1154 ggaattgtacttctgcga 1174
DB 1435 AAATTTCTACTCTTAGCA 1455

RESULT 10
US-09-014-969-14/c
; Sequence 14, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallee, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014.969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-014-969-14

```

```

Query Match
Best Local Similarity 52.5%; Pred. No. 0.022;
Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```

```

QY 1028 ttaattcagagattttcttccagtttagtcctgtattttatgtatattcttacc 1087
Db 2385 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 2326
QY 1088 ttgatttgatgatttttttccctaaatttataatttccctaattctgtaagtaa 1147
Db 2325 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 2266
QY 1148 tgaatgagatattgtactttctgtcaataatagacagacatcgcaaaaaaa 1207
Db 2265 ttttttttttttttttttttttcccaaaaggttattgtcacatttaagracaaaat 2206
QY 1208 aaa 1210
Db 2205 caa 2203

```

```

RESULT 11
US-08-157-101A-4/c
; Sequence 4, Application US/08157101A
; Patent No. 56808032
; GENERAL INFORMATION:
; APPLICANT: KURIHARA, TATSUYA
; APPLICANT: MATSUKURA, SHIGEKAZU
; APPLICANT: TSUBOKA, NOBEO
; APPLICANT: ARIMA, KENJI
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION
; TITLE OF INVENTION: PLASMIDS THEREFOR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/157,101A
; FILING DATE: 05-APR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: TITUS, MARILAN K
; REGISTRATION NUMBER: 35843
; REFERENCE/DOCKET NUMBER: 9437/204199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUCH

```

```

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1066 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-157-101A-4

```

```

Query Match
Best Local Similarity 56.2%; Pred. No. 0.018;
Matches 82; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```

```

QY 1041 tttcttcagattagtcctgtattttatgtatatttccctattgtatgtat 1100
Db 1037 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 978
QY 1101 gatttttccctaaatttataatttccctaattctgtaagttaattgagatatt 1160
Db 977 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 918
QY 1161 tgaacttctgtcaataatagacaa 1186
Db 947 ttcatcttcctcccaacatttagcatma 892

```

```

RESULT 12
US-09-370-253-1/c
; Sequence 1, Application US/09370253
; Patent No. 6165792
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Sakai, Hajime
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Amino Acid Transporters
; FILE REFERENCE: Bb-1200
; CURRENT APPLICATION NUMBER: US/09/370,253
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: 60/097,222
; EARLIER FILING DATE: August 20, 1998
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 1882
; TYPE: DNA
; ORGANISM: Hordeum vulgare
; US-09-370-253-1

```

```

Query Match
Best Local Similarity 56.2%; Pred. No. 0.022;
Matches 82; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```

```

QY 1041 tttcttcagattagtcctgtattttatgtatatttccctattgtatgtat 1100
Db 1880 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 1821
QY 1101 gatttttccctaaatttataatttccctaattctgtaagttaattgagatatt 1160
Db 1820 ttttttttttttttttttttttttttttttttttttttttttttttttttttt 1761
QY 1161 tgaacttctgtcaataatagacaa 1186
Db 1760 ttccggaatattctatgcacagacaa 1735

```

```

RESULT 13
US-09-247-373B-33/c
; Sequence 33, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN

```

| | | | | |
|-----------------------|--------------|------------------|----------------|-------------|
| Query Match | 3.68; | Score 43.2; | DB 4; | length 117; |
| Best Local Similarity | 57.48; | Pred. No. 0.023; | | |
| Matches 78; | Conservative | 0; | Mismatches 58; | Indels 0; |
| | | | Gaps | 0; |

RESULT 14
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654

```

?
?
? TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
?
? FILE REFERENCE: 09404/052001
?
? CURRENT APPLICATION NUMBER: US/09/128,155
?
? EARLIER APPLICATION NUMBER: US 60/091,650
?
? EARLIER FILING DATE: 1998-07-02
?
? EARLIER APPLICATION NUMBER: US 60/054,646
?
? EARLIER FILING DATE: 1997-08-04
?
? NUMBER OF SEQ ID NOS: 18
?
? SOFTWARE: FastSeq for Windows Version 3.0
?
? SEQ ID NO 17
?
? LENGTH: 176373
?
? TYPE: DNA
?
? ORGANISM: Homo sapiens
?
? FEATURE:
?
? NAME/KEY: misc-feature
?
? LOCATION: (1)...(176373)
?
? OTHER INFORMATION: n = A,T,C or G
?
? US-09-128-155-17

```

| | | | | |
|-----------------------|--------|-----------------|-------|---------------|
| Query Match | 3.68; | Score 43; | DB 3; | Length 176373 |
| Best Local Similarity | 48.2%; | Pred. No. 0.19; | | |

RESULT 15
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5280526

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 20:24:37 ; Search time 1589.3 Seconds

(without alignments)
10275.792 Million cell updates/sec

Title: US-09-805-919-3

Perfect score: 1210

Sequence: 1 tatattcaatgaaacatg.....ttcgcaaaaaaaaaaaaaa 1210

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: EST

1: em_estba:*
2: em_estbun:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_esti:*
9: gb_esti:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 349 | 28.8 | 639 | 10 | BI434492 EST537253 |
| 2 | 335.4 | 27.7 | 538 | 9 | AT775018 EST256118 |
| 3 | 301.6 | 24.9 | 533 | 9 | AT697599 EST267042 |
| 4 | 238.4 | 19.7 | 432 | 9 | AI490332 EST248658 |
| 5 | 70 | 5.8 | 654 | 10 | BF645984 NF065E11E |
| 6 | 69.8 | 5.8 | 498 | 10 | BF633219 NF054H02D |
| 7 | 67.8 | 5.6 | 603 | 9 | AM695109 NF091D01S |
| 8 | 67.6 | 5.6 | 329 | 9 | AL513719 NF0513719 |
| 9 | 65.8 | 5.4 | 1045 | 12 | CNS033YE4 |
| 10 | 65 | 5.4 | 625 | 10 | BF637868 |
| 11 | 65 | 5.4 | 635 | 10 | BF637888 NF030B07P |
| 12 | 64.8 | 5.4 | 945 | 10 | BF635609 NF061D05P |
| 13 | 63.6 | 5.3 | 638 | 9 | BM358135 GA_Ea000 |
| 14 | 62.6 | 5.3 | 991 | 12 | AL513901 AL513901 |
| 15 | 62.6 | 5.2 | 353 | 9 | AL074343 Drosophila |
| 16 | 62.6 | 5.2 | 503 | 9 | AL515235 AL515235 |
| 17 | 62.6 | 5.2 | 546 | 10 | AL513809 AL513809 |
| | | | | | BF643246 NF003C11E |

| | | | | | |
|----|------|-----|------|----|----------|
| 18 | 62.4 | 5.2 | 660 | 10 | BS591192 |
| 19 | 62.4 | 5.2 | 947 | 12 | CNS015IM |
| 20 | 61.4 | 5.1 | 379 | 9 | AL514359 |
| 21 | 61 | 5.0 | 425 | 9 | AL514791 |
| 22 | 60.8 | 5.0 | 633 | 9 | AL513979 |
| 23 | 60.6 | 5.0 | 365 | 9 | AL515373 |
| 24 | 60.4 | 5.0 | 329 | 9 | AL514627 |
| 25 | 60 | 5.0 | 1043 | 12 | CNS0145P |
| 26 | 59.6 | 4.9 | 319 | 9 | AL513877 |
| 27 | 59.6 | 4.9 | 951 | 12 | A2672893 |
| 28 | 59.2 | 4.9 | 344 | 9 | AL513975 |
| 29 | 59.2 | 4.9 | 894 | 12 | BH137168 |
| 30 | 59 | 4.9 | 523 | 9 | AL514015 |
| 31 | 58.8 | 4.9 | 712 | 9 | BE038709 |
| 32 | 58.4 | 4.8 | 1101 | 12 | CNS003DQ |
| 33 | 58.2 | 4.8 | 309 | 9 | AI490927 |
| 34 | 58.2 | 4.8 | 468 | 9 | AI498859 |
| 35 | 58.2 | 4.8 | 470 | 9 | AI490270 |
| 36 | 58.2 | 4.8 | 886 | 12 | BH177277 |
| 37 | 58.2 | 4.8 | 886 | 12 | CNS07JDX |
| 38 | 58.2 | 4.8 | 941 | 10 | BM415213 |
| 39 | 58 | 4.8 | 344 | 9 | AL513875 |
| 40 | 58 | 4.8 | 363 | 9 | AL514473 |
| 41 | 57.6 | 4.8 | 956 | 12 | CNS04MDT |
| 42 | 57.6 | 4.8 | 392 | 9 | AL514511 |
| 43 | 57.6 | 4.8 | 449 | 9 | AL513999 |
| 44 | 57.6 | 4.8 | 920 | 12 | CNS0062R |
| 45 | 57.6 | 4.8 | 954 | 12 | BH162327 |

ALIGNMENTS

RESULT 1
LOCUS BI434492 639 bp mRNA linear EST 21-AUG-2001
DEFINITION EST537253 P. infestans-challenged leaf Solanum tuberosum cDNA clone
PCBR09 5' sequence, mRNA sequence.
ACCESSION BI434492
VERSION BI434492.1 GI:15259182
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 639)
Austrop, S., Griffiths, H.M., Smart, C.D., Cho, D., Chiemingo, A., Bougri, O., Buell, C.R., Ronning, C.M., Fry, W.E. and Baker, B. Generation of ESTs from Potato Leaves Challenged with Phytophthora infestans, Compatible Interaction
Unpublished (2000)
JOURNAL Contact: Cathy Ronning
COMMENT The Institute for Genomic Research
For clone info: Please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdnaregen.com
Seq primer: MJ3F-R.

FEATURES

Source

Location/Qualifiers
1..639

/organism="Solanum tuberosum"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="PCBR09"

/clone_1lb="P. infestans-challenged leaf"

/tissue_type="leaf"

/dev_stage="6 week old"

/lab_host="SOLR"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Whole plants were challenged with 450,000 sporangia/ml P. infestans (US 940501) in bloton (Madison, Wisconsin). Leaf tissue was collected at 1, 2, 5, 12, and 24 hours post-challenge and frozen in liquid nitrogen immediately upon removal. Kennebec plants showed

no signs of HR. Katahdin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."

BASE COUNT 200 a 144 c 101 g 194 t
ORIGIN

Query Match 28.8%; Score 349; DB 10; Length 639;
Best Local Similarity 76.9%; Pred. No. 3,1e-48;
Matches 489; Conservative 0; Mismatches 125; Indels 21; Gaps 4;

QY 1 tatattcaattgaaacatggttgcacaaaggagctcttctgctgtgtttgttactg 60
DB 9 TATATTCAATTGACAAAATGTCATTAACAATTCATTTTTCCTAGTGTTCCTCTG 68
QY 61 tgaactcataaagcagttctaaagcgagagatatt--tctaatgtctccctctct 117
DB 69 TTTCATTCATACAGAGCTGTATAGCCCAATATAGTACGTCTCATGTGTCTCTTCT 128
QY 118 ggggtgtatt---caataaattctccctccagctgagctgattcccgagctgtg 174
DB 129 GTGGCGATATTCGCCATATAAATTCCTTCGATGAGACTGATCCTCGAGCATTTG 188
QY 175 gtgagcgagatatagctgagctgagctgagacacacgctgttcaataacaatcca 234
DB 189 GCAACATGATAGCAACTGATGCGCAGAACATCAACCTTCTACACTTCAAGTCA 248
QY 235 gaatttcgacgtacaggaatcaactacagaagctactcaataaggctactgtactg 294
DB 249 GAAATTTCTATCGCAGGAAATTAATCAACAGCTACATGATGAGGCTACTCGATCTG 308
QY 295 gcttaaatgacagagaagaattgacaggttttccaataacagaggaagtatt--g 351
DB 309 GCCTAAATGATCAGAAAGAAATGCTGTGTTTCTCCTGATTAACAGCAATTTTACG 368
QY 352 atgcgatgactgacaaatcttgaatgggttcgtgttaacaagatatacaactgtgca 411
DB 369 GAGGCTTAATGATGAATTTTCCATTAATTCATTAATTAACATTAAGCAATCACTATGCA 428
QY 412 actgtcagctcctatcaaatctgacagtatatctcacaagttttgttaacaaatt 471
DB 429 ACHTGTCACTCCTATCAACTGCTCAATACATTCACACGACTTTTGTAGAAACATA 488
QY 472 caa-----cgggttttagctacctgtgcatagaagaatatgtcaagcttcg 519
DB 489 CAATCTGCCCCAAGCCCAATTTAGTTATCTGTCTTAAGAAATATGCGAAGCTCGG 548
QY 520 attgtgctgagcgctgtaaggttgaactgttgcatggtcctctctcagaagcatctat 579
DB 549 ATTGTGGAACGGTTGCAAGGTGAAACGTGGCGGTGCTTCACTCCAGGTATTTTGA 608
QY 580 caaacaagtcgtctacgttatacaagcacaca 610
DB 609 CAAACAAGTCGCTTTGTTAGGAAGCATCA 639

RESULT 2
LOCUS A1775018 538 bp mRNA linear EST 18-MAY-2001
DEFINITION EST256118 tomato resistant, Cornell Lycopersicon esculentum cDNA
ACCESSION A1775018
VERSION A1775018.1 GI:5273059
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 538)

AUTHORS D' Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,
Liang, F., Upton, J., Romling, C.M., Craven, M.B., Fujii, C.Y., Bowman,
C.U., Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,
Giovannoni, J.J., and Martin, G.B.
Generation of ESTs from Pseudomonas resistant tomato
Unpublished (1999)
CONTACT: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source Location/Qualifiers
1..538
/organism="Lycopersicon esculentum"
/cultivar="RI1-12 (35S::Pto in Rio Grande x Money Maker)"
/db_xref="taxon:4081"
/clone="CLER1435"
/clone_lib="tomato resistant, Cornell"
/tissue_type="leaf"
/dev_stage="4-week old"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; CLER - Tomato Pseudomonas resistant EST library.
Directionally cloned cDNAs inserted into pBluescript SK(-
) at 5' end with EcoRI and 3' end with XhoI site."

BASE COUNT 145 a 126 c 107 g 160 t
ORIGIN

Query Match 27.7%; Score 335.4; DB 9; Length 538;
Best Local Similarity 79.1%; Pred. No. 5.8e-46;
Matches 425; Conservative 0; Mismatches 106; Indels 6; Gaps 2;

QY 95 ttctcaatggttccctctctctctggtgatat--caataaattccctccgact 151
DB 1 TTCTCATGTGTCCCTTCATCTGTGGCGATGAGCAACTTAATTTCTTCCGATT 60
QY 152 gaggaactgataccagacatgtgtgtagacgagatatgactcgatccagaacaaca 211
DB 61 GAGGTGTGATCCCGACAGCATTTGTAGACGTGATATGACATCGATGCCAGATTAAGA 120
QY 212 aaccggttcaataataatccaagaatttcgaagtaagaagaattactcagaagcta 271
DB 121 AACCATTTTCCATTAACAATGCGGTAAATTCATCTACGAGAGATTAATACCACTT 180
QY 272 ctcaataaggtactgtactcgtgacctaaatgatcagaagaataatgcacagttttcc 331
DB 181 CTCATACAGGCTACGTGATCCGAGCTGAAATACAGCCGAAATTCCTGTTTTCC 240
QY 332 aatatacagggcaagtatagagcactgactgacaaatcttgatggtgtgtgttaa 391
DB 241 ACAACACAGGGCAAGTTAGAGTCCATGACCTTAATCTTGGATGGCTTCGTTGGA 300
QY 392 caatgataloactatgtaactgtctactcactcaaltgcatacagatlatatctac 451
DB 301 TAAAGATTAATATATGTCACTGTGCGGTAATCAACCTCCAGATTAATTCCTAC 360
QY 452 aagtttttgaagaaaaattcaa--cgggttttagctacctgtgcatagaagaattt 508
DB 361 AAGTTTTTGTAGCAGCATCAAAATTCAGTTTGTAGCTATCTTGTGAGGTGAATTT 420
QY 509 gcaagctcgagatttgcgtgagcgctgtaggttgaactgttgatagctcctgctcc 568
DB 421 GCAAGCTTCAGATTAGCAGTGGCGGTGCAAGATTGAACCTGCGATGAGAGCGCTCC 480
QY 569 aggcatttcatcaacaagtctgctacgttatacaagcacacataagcctggtcta 625
DB 481 GGGGATTTCAACCTAAGTATCGTTCCTGTTATCCAGCATTCACCAAGCTCGGGTTA 537

RESULT 3
A1897599

```
LOCUS      A1897599      533 bp      mRNA      linear      EST 18-MAY-2001
DEFINITION ESM267042 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
ACCESSION  AF897599
VERSION     A1897599.1  GI:5603501
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
            Asteridae: euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE  1 (bases 1 to 533)
AUTHORS   Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,
            Liang, F., Upton, J., Romning, C.M., Craven, M.B., Fujii, C.Y., Bowman
            , C.L., Niernan, M., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley
            , S.D. and Giovannoni, J.
            Generation of ESTs from tomato carpel tissue
            Unpublished (1999)
TITLE      JOURNAL
COMMENT    Contact: CUGI
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.
FEATURES   Location/Qualifiers
            source
            1..533
            /organism="Lycopersicon esculentum"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /clone="cLED28G17"
            /clone_1db="tomato ovary, TAMU"
            /tissue_type="carpel"
            /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
            /lab_host="X11-Blue MRP"
            /note="Vector: Bluescript SK(-); Site.1: EcoRI; Site.2:
            XhoI; cLED - Tomato Carpel EST library. OligodT-primed and
            directionally cloned cDNA in vector lambda ZAP II with 5'
            and 3' ends located at the EcoRI and XhoI sites,
            respectively."
BASE COUNT 135 a      103 c      120 g      175 t
ORIGIN
Query Match      24.9%; Score 301.6; DB 9; Length 533;
Best Local Similarity 75.3%; Pred. No. 2,1e-40;
Matches 405; Conservative 0; Mismatches 124; Indels 9; Gaps 2;

QY 305 tcagagagaaatgacagttttccaatccacagggaagtatagatgcagtag 364
DB 1 TCAGACCGAAATTGCTGTTTCCACACAGGCGCAAGTTACGATGCCATGACTAG 60
QY 365 ccaactcttgaatgggtcgtgtaacaatgatatcaactaigtcaactgctcctcc 424
DB 61 CCTATCTTCGGATGCGTTCGTGTGATTAAGATATCATATATCTCACTGTCGGCTCA 120
QY 425 tatcaatcgttcacagtatctctcaagaattttgtagaacaaatcaa---cgggtt 481
DB 121 AATCAACTCTCAGATATCATCTCTACAGGTTTGTAGACACACTCAAAATTCITAGTTT 180
QY 482 tagctacttgatcataaagaataatgcaagcttgagatttgctggcgctgtagt 541
DB 181 TAGCTATCTTGTCGAGTGAAATATTCACAGCTTCAGATTTAGCACTCGGTTGCAGAGT 240
QY 542 tgaacgtgtgcatggtcctctgctccaggcattcaacaagaatgctctacgttacc 601
DB 241 TGAACGTGTGCGATGATGATCTCCGGGATTCACACATACGTATGCTTCGTCATC 300
QY 602 aagcaacatcaagcgccggtttaggtttgagcttcttggaaggtatctggtatg 661
DB 301 CAGCAATTCACCAAGCTCGGGTTACGGATTTGATCTCTCTTGGAGGCGGATTTCTTATG 360
QY 662 tagaaattgcagccggagctgtggtgggtgagtgcatatigaaagaaacagcagcagc 721

RESULT 4
A1490332
LOCUS      A1490332      432 bp      mRNA      linear      EST 18-MAY-2001
DEFINITION ESM248658 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
ACCESSION  A1490332
VERSION     A1490332.1  GI:4365642
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
            Asteridae: euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE  1 (bases 1 to 432)
AUTHORS   Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,
            Liang, F., Upton, J., Romning, C.M., Craven, M.B., Fujii, C.Y., Bowman
            , C.L., Niernan, M., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley
            , S.D. and Giovannoni, J.
            Generation of ESTs from tomato carpel tissue
            Unpublished (1999)
TITLE      JOURNAL
COMMENT    Contact: CUGI
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            Location/Qualifiers
            source
            1..432
            /organism="Lycopersicon esculentum"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /clone="cLED24P14"
            /clone_1db="tomato ovary, TAMU"
            /tissue_type="carpel"
            /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
            /lab_host="X11-Blue MRP"
            /note="Vector: Bluescript SK(-); Site.1: EcoRI; Site.2:
            XhoI; cLED - Tomato Carpel EST library. OligodT-primed and
            directionally cloned cDNA in vector lambda ZAP II with 5'
            and 3' ends located at the EcoRI and XhoI sites,
            respectively."
BASE COUNT 112 a      85 c      99 g      136 t
ORIGIN
Query Match      19.7%; Score 238.4; DB 9; Length 432;
Best Local Similarity 74.9%; Pred. No. 5,8e-30;
Matches 328; Conservative 0; Mismatches 101; Indels 9; Gaps 2;

QY 305 tcagagagaaatgacagttttccaatccacagggaagtatagatgcagtag 364
DB 1 TCAGACCGAAATTGCTGTTTCCACACAGGCGCAAGTTACGATGCCATGACTAG 60
QY 365 ccaactcttgaatgggtcgtgtaacaatgatatcaactaigtcaactgctcctcc 424
DB 61 CCTATCTTCGGATGCGTTCGTGTGATTAAGATATCATATATGTCATAGTCGCGGTCA 120
QY 425 tatcaatcgttcacagtatctctcaagaattttgtagaacaaatcaa---cgggtt 481
DB 121 AATCAACTCTCAGATATCATCTCTACAGGTTTGTAGACACAGCTCAAAATTCITAGTTT 180
```


| | | | |
|----|-----|--|-----|
| QY | 482 | tagtacctgtcataaagaagaatatggaaccttcggaattgctgtgcggctgaaggt | 541 |
| Db | 181 | TAGTACTCTTGTGCTGAGGTGAATATTGGAAAGCTTCAGATTAGGAGTCGGTTGCAGGT | 240 |
| QY | 542 | tgaactgttgcatagtgttcctctgtctccaaagcatttcataaacaagtgcgtacgtatc | 601 |
| Db | 241 | TGAACCTGTGGCATGGAGTGACACCTCCGGGGATTTCACCTACGATATGCTTCGTATC | 300 |
| QY | 602 | aagacacatacaaggcctgtgcttatgtgtttaaagcttcttggaaagtaactcgttatg | 661 |
| Db | 301 | CAGATTTCACCAAGCTCTGGGTTACGGGATTTAATCTCTTGGAAAGCGGATTTCTATG | 360 |
| QY | 662 | tagaaatctgcgaacggagatctgtgggggttagtgcacatacttgaagaaacagcgcagaagc | 721 |
| Db | 361 | CAGAGAGTGTG-----GGAGAGATGATGATATGTGTTTCAAAGACANATAGGAGATGTAGC | 414 |
| QY | 722 | tactgtcgttattgtgtg | 739 |
| Db | 415 | AACCTTGATTCATTACTG | 432 |

| RESULT | 5 |
|------------|--|
| LOCUS | BF645984 |
| DEFINITION | BF645984 654 bp mRNA, linear EST 20-DEC-2000 NF065E11EC1F1086 Elicited cell culture Medicago truncatula CDNA clone NF065E11EC 5', mRNA sequence. |

| | |
|-----------|--------------------------------------|
| ACCESSION | BF645984 |
| VERSION | BF645984.1 |
| KEYWORDS | GI:11911113 |
| SOURCE | EST. |
| ORGANISM | barrel medic. Medicago truncatula |

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

REFERENCE
1 (bases 1 to 654)
Torres-Hernandez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
AUTHORS Flores-Hernandez, R., Imman, U.T., Weller, J.W. and May, G.D.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
JOURNAL Center for Medicago Genomics Research
Unpublished (2000)

COMMENT
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 654 Std Error: 0.00
Plate: 065 row: E column: 11
Seq primer: TCACACAGGAAACAGCTGTGAC.

```

FEATURES
source
Location/Qualifiers
1. 654
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NP065818C"
/clone_lib="Elicited cell culture"
/tissue_type="Cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"
/note="Vector: Lambda zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5, 1, 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."
BASE COUNT
214 a 145 c 102 g 192 t 1 others
ORIGIN

```

| | | | | |
|-----------------------|--------------|------------------|----------------|-------------|
| Query Match | 5.8%; | Score 70; | DB 10; | Length 654; |
| Best Local Similarity | 62.1%; | Pred. No. 0.026; | | |
| Matches 128; | Conservative | 0; | Mismatches 75; | Indels 3; |
| | | | Gaps | 1 |

| | | | | |
|----|-----|--|----------------------------|-----|
| QY | 102 | tgtgtccctctctctcgtggtgatat--- | ccaatnaattcccttcgactgagact | 158 |
| | | | | |
| Db | 165 | tgtcccccttttcattgtggcgaataatgcaaacattaaagtatcccttcgcattgaagat | 224 | |
| | | | | |
| QY | 159 | gattcccgagcatctgtgtgtagacgsggatalgaactcgatctgcacgaacacaaacgctg | 218 | |
| | | | | |
| Db | 225 | gattccgacaaactgggtggatctccaggttagagattttcctcggaataaaacttcacatg | 284 | |
| | | | | |
| QY | 219 | tccaattacaatccagaatttttcgacgtacaggaatttaactacaagaactactcaata | 278 | |
| | | | | |
| Db | 285 | tttaaacctccttttcgggttaattaccatgataaaatcaatcaatcacaatracacacatt | 344 | |
| | | | | |
| QY | 279 | aggctactgtacccgctgactaatga | 304 | |
| | | | | |
| Db | 345 | cggttagttgattcctggaaattaaaga | 370 | |
| | | | | |

| | | | | | |
|------------|--------------------------|---------------------|------------|------------|-----------------|
| RESULT | 6 | | | | |
| LOCUS | BF633219 | | | | |
| DEFINITION | BF633219 | 498 bp | mRNA | linear | EST 19-DEC-2000 |
| | NF054H02DT1F1026 Drought | Medicago truncatula | cDNA clone | NF054H02DT | |
| ACCESSION | 5'', mRNA sequence. | | | | |
| VERSION | BF633219 | | | | |
| KEYWORDS | BF633219.1 GI:11897377 | | | | |
| SOURCE | EST. | | | | |
| | Bartel medic. | | | | |

ORGANISM
Medicago truncatula
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
Rosidae: eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|--------------------|---|--|--------------------|--|
| 1 (pages 1 to 498) | Torrez-Herez, I., Scott, A.D., Harris A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D. | Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula drought library | Unpublished (2000) | Contact: May GD Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA |

```

tel: 580 221 7391
fax: 580 221 7380
Email: gdmayenoble.org
Insert Length: 498 Std Error: 0.00
Plate: 054 row: H column: 02
Seq primer: TCACACAGGAAACACGCTGAC.
Location/Qualifiers
1..498
/organism="Medicago truncatula"
/db_xref="taxon:3880"

```

```

/clone="NF054H02DT"
/clone_lib="Drought"
/tissue_type="Plantlets"
/dev_stage="Pooled timepoints"
/notes="Vector: lambda zap; Contains a mixture of entire
plantlets harvested in a series of days post-watering
timepoints."
BASE COUNT      167 a      110 c      75 g      146 t
ORIGIN
Query Match      5.8%; Score 69.8; DB 10; Length 498;
Best Local Similarity 59.2%; Pred. No. 0.031;
Matches 138; Conservative 0; Mismatches 92; Indels 3; Gaps 1
QY 75 agtctaagcgacagatattctcaaatgtgcacctcttcctcgagtgatat--tcaa 131

```

QY 132 ataaattcccttcgcgactgaggaactgatcccagacatgttctgttagacgcgatatgag 191
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 192 cgtgatgtgcagagaaacaacccaagtcgttcaatcaacaacccgaatttcgacgaaag 251
 Db 289 TTATCTCTGGGAATACACATCTCACACCGCTTAACCTCTGTTTATGAGTAAATCTATGTCGACG 348
 QY 252 gaattactacagaagctactcaataagctacttgatccctgcgcctaaatga 304
 Db 349 GAATATACCTACATAAAGTACACAACTTGTGTTAGTATGATCCAGGAATGTAGA 401

| | | | | | |
|--|------------|--|--------|-------|--------|
| | RESULT | 7 | | | |
| | AM695109 | | | | |
| | LOCUS | AM695109 | 603 bp | -mRNA | linear |
| | DEFINITION | NF09JID01STIF1012 Developing stem Medicago truncatula cDNA clone | | | |
| | ACCESSION | NF09JID01ST 5', mRNA sequence. | | | |
| | VERSION | AM695109 | | | |
| | KEYWORDS | AM695109.2 GI:11935567 EST. | | | |
| | SOURCE | bareil medic. | | | |
| | ORGANISM | Medicago truncatula | | | |

REFERENCE
1 (bases 1 to 603)
AUTHORS
He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell

| | |
|---------|--|
| TITLE | Expressed Sequence Tags from the Samuel Roberts Noble Foundation |
| JOURNAL | Medicago truncatula stem library |
| COMMENT | Unpublished (2000) On Apr 14, 2000 this sequence version replaced g1:7568871. |

Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel.: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 650 Std Error: 0.00
Plate: 091 Row: D Column: 01
Seed primer: TCACACAGGAAACGCAATAC.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1: .603 |

```

/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_id="MF091D01S1"
/clone_lib="Developing stem"
/tissue_type="stem"
/dev_stage="Pooled developmental"
/notes="Vector: lambda zap; Contains a mixture of
intermodal stem segments"
191 a
122 c
109 g
181 t

```

| | | | | |
|---------------------------|--------|------------------|-----------|-------------|
| Query Match | 5.6%; | Score 67.8; | DB 9; | Length 603; |
| Best Local Similarity | 61.7%; | Pred. No. 0.062; | | |
| Matches 108; Conservative | 0; | Mismatches 67; | Indels 0; | Gaps 0 |

Qy 130 aacaaatcttccttcgcagctagagcagcagcagatcgggtatgacagcgatattg 189
Db 176 ACATPAAACATCTCTTTCGATTCGAGAAAGATCCACCAAACTCGCGTATCCGAGTATG 235
Qy 180 agctcagctgcagaaacaaacacgcgtygtcaatcacaaatccagaatttcgaaglac 249
Db 236 AGTTATCTCTGGAAAAACATCCNCAACGTTAACTTTTCACGTAACATCATGTGA 295
Qy 250 aggaatctaactacagaagctcactcattcaatagagctattatctcgtgctaataga 304
Db 296 AATCATTAATTAATCGAAACTCATCCATCATGAGATTAATTGGAATTCAAG 350

| | | | |
|--|------------|--------------------------|---|
| | RESULT | 8 | |
| | AL517719/c | | |
| | LOCUS | | |
| | DEFINITION | Al517719 | 329 bp mRNA; linear EST 13-FEB-2001 |
| | | ALI517719.LTI.NFLOO6.PL2 | Homo sapiens cDNA clone GLOBAR007ZH01 3 |
| | | prime, mRNA sequence. | |
| | ACCESSION | Al517719 | |
| | VERSION | Al517719.1 | GI:12777213 |
| | KEYWORDS | EST. | |
| | SOURCE | human. | |

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 329)
AUTHORS Li, W. B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqinfo@genoscope.cns.fr, Web : www.genoscope.cns.fr.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .329 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBA007ZH01"
/clone_lib="L11_NFL006_PL2"
/issue_type="Placenta"
/vector="pcmwsv0r1 6; Site_1: NotI; 1st strand cDNA
/Note="Vector: pcmwsv0r1 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-Oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pcmwsv0r1 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@litech.com URL :
http://fulllength.invitrogen.com"

```

| | | | | |
|--------------------------|--------|------------------|-----------|-------------|
| Query Match | 5.6%; | Score 67.6; | DB 9; | Length 329; |
| Best Local Similarity | 42.9%; | Pred. No. 0.085; | | |
| Matches 73; Conservative | 44; | Mismatches 53; | Indels 0; | Gaps 0 |

1041 tttttctttccagatttaotcctctatatttatttgaatattcttaacctaattgaattgatat 1100

[illegible][illegible]

```

1101 gatctttttccctaaatttcatacaatttcctgaagaaatcggaatgacatc 1100
||||| ::: :|::| :||| :::| :|:: :|:: :|:: :|::

```

[illegible]

1161 tgtaactttctgtcataataatagaacaagacattcgcaaaaaaaaaaaaaa 1210

b
81 KTTTWWWWWNAWAAAIAWAAAAAIAAAAWWAFAAAAAAAAAA 32

| LOCUS | DEFINITION | SEQUENCE | 1045 bp | DNA | linear | GSS 18-MAY-2000 |
|----------|---|----------|---------|-----|--------|-----------------|
| CNS03YE4 | Tetradon nigroviridis genome survey sequence T7 end of clone 066003 of library G from Tetradon nigroviridis, genomic survey | | | | | |

ACCESSION AL266197.1 GI:7987962
VERSION
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradodon nigroviridis.
ORGANISM Tetradodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neuteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetradodon

REFERENCE 1 (bases 1 to 1045)
 AUTHORS Roest-Crolius,H., Jallion,O., Dasilva,C., Fizesnes,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetler,F., Saulin,W., Bernot,A. and
 Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1045)
 AUTHORS Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizesnes,C., Wincker,P., Brothier,P., Quetler,F.,
 Saulin,W. and Weissenbach,J.
 TITLE Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1045)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/tetraodon.
 FEATURES
 Location/Qualifiers
 source 1..1045
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="068003"
 /clone_1lb="G"
 /note="Genoscope sequence ID : COB068AH02LP2-end : T7"
 BASE COUNT 374 a 155 c 166 g 298 t 52 others
 ORIGIN
 Query Match 5.4% Score 65.8; DB 12; Length 1045;
 Best Local Similarity 50.3% Pred. No. 0.11;
 Matches 92; Conservative 24; Mismatches 67; Indels 0; Gaps 0;
 QY 1028 ttaactagagattctctccctcagttagctcgtatttattgatacttacta 1087
 DB 955 ttttttttttatttttttttttttttttttttttttttttttttttttt 896
 QY 1088 ttgtattgtatgattttttcttcaaatctataattcttcaattcttgaagta 1147
 DB 895 ttt 836
 QY 1148 ttgaatgagatttcttctcgtcaataagacagacattcgcaaaaaaa 1207
 DB 835 ttatamwmttt 776
 QY 1208 aaa 1210
 DB 775 AAA 773
 RESULT 10
 LOCUS BF637888 625 bp mRNA linear EST 19-DEC-2000
 DEFINITION NF030807P1.1F1059 phosphate starved leaf Medicago truncatula CDNA
 ACCESSION BF637888
 VERSION BF637888.1 GI:11902046
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 REFERENCE 1 (bases 1 to 625)
 AUTHORS Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
 ,H.R., Imman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 Medicago truncatula phosphate-starved leaf library

JOURNAL Unpublished (2000)
 COMMENT Contact: Harrison MJ
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA
 Tel: 580 221 7325
 Fax: 580 221 7380
 Email: mjharrison@noble.org
 Insert Length: 625 Std Error: 0.00
 Plate: 030 Row: B Column: 07
 Seq primer: TCACACGAAACACGCTATGAC.
 FEATURES
 source 1..625
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone="NF030807PL"
 /clone_1lb="Phosphate starved leaf"
 /tissue_type="leaf"
 /dev_stage="trifoliolate"
 /note="Vector: Lambda Zap; At the trifoliolate stage, M.
 truncatula plants were transplanted to phosphate-free sand
 and grown for a further 30 days. During this 30 day
 period, the plants were fertilized twice weekly with 1/2
 Hoaglands solution containing only 20uM potassium
 phosphate. RNA was prepared from above ground tissues."
 BASE COUNT 206 a 134 c 101 g 184 t
 ORIGIN
 Query Match 5.4% Score 65; DB 10; Length 625;
 Best Local Similarity 57.9% Pred. No. 0.18;
 Matches 135; Conservative 0; Mismatches 95; Indels 3; Gaps 1;
 QY 75 agttcaagcgcagagatattcctaagtgtccctctcctcgcgtgatattc 131
 DB 172 ATTACTATATACAAATTAACCAACATATTGTCACACTTCTCATGCGGCAAAATCACAAAC 231
 QY 132 ataaattcctccctcgcagcgtgagatccgagacattgtgtgagcgcgattgag 191
 DB 232 ATATCACATCCTTTTCGCTTATGAATGACCAACAAAGCTGTGATCCGAAATGCGAG 291
 QY 192 ctgattgcgcagacaacaaacggtgttaattacaacaaatttcgcagtaag 251
 DB 292 TTATCCCTCGAAACAAACATACCGCTTATACCTGTGTTTCAGAGTAATACATATGCGAG 351
 QY 252 gaatttaactcagaagctactcaataagacttgaacctgagcctaata 304
 DB 352 GAAATCAACATCGTAATTAATACCAATTCGGCTAGTGTATCCAGTAATTGAAGA 404
 RESULT 11
 LOCUS BG455609 655 bp mRNA linear EST 19-MAR-2001
 DEFINITION NF061D05P1.1F1044 phosphate starved leaf Medicago truncatula CDNA
 ACCESSION BG455609
 VERSION BG455609.1 GI:1378934
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 REFERENCE 1 (bases 1 to 655)
 AUTHORS Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
 ,H.R., Imman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 Medicago truncatula phosphate-starved leaf library
 JOURNAL Unpublished (2000)
 COMMENT Contact: Harrison MJ
 Plant Biology Division
 The Samuel Roberts Noble Foundation

Best Local Similarity 54.4%; Pred. No. 0.55; Matches 92; Conservative 14; Mismatches 63; Indels 0; Gaps 0;

```

QY 1042 ttcttcacagtttaagtcctgfatlttattgatacttaccatttgatgcgtalg 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1102 atttttcttaaaatttataaatttcctaattccttgtaagtaattgaatgataatt 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1162 gtacttcctgcaataaagaacagcattcgcaaaaaa 1210
    : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: June 4, 2002, 20:54:27
Job time: 1790 sec

| | | |
|----|---|-----|
| 51 | CATACACACTTCTTAAGCGCAGAGATATTCTTCATATGTCCTCTTCT | 100 |
| 52 | YSGIYASPILEGLINILEYSPHEPROHEARGLEUATHRRASPROGLU | 50 |
| 53 | GGGATATTTCTAAATAATAATTTCCCTTCGACGAGAGATGTATCCGAG | 150 |
| 54 | HSACYSGLIYARGARGILTYTGLULLEUASPQSGILNASNANDINTRA | 67 |
| 55 | CATTGGTAGAGACGGGATATGACGTGGATGGCAGAACACCAACCTG | 200 |
| 56 | LPHEASNYRILYSSERARGILPEHEAPVALINGLULEASNYRARGS | 84 |
| 57 | GTTCAATTTCANATCCAGANATTTTGAGTCAGCGGAATTAATTAAGAA | 250 |
| 58 | ERTYSERTLEARGLEULNUASPROGLYUENASNAPGLNARGLUASN | 100 |
| 59 | GCTACTCATTAAGGCTACTTATCTGCTGGCCATAATGATCAGANAGAAAT | 300 |
| 60 | CYSTHVALPHEPROASNHARGALASERTYRASPPLAMETHSERGL | 117 |
| 61 | TGCACAGTTTTCCAAATCACAGGCGCAGATTATGATGCCATGACTAGCCA | 350 |
| 62 | NILEPHEGLUTRPPVALARGVALASNASNAPILEASNYRVALASNYCAL | 134 |
| 63 | AACTTTGAAATGGGTCTGTTTAACATGAAATCACTAATGTCACACTGC | 400 |
| 64 | EUALAPROLLEASNSESRGINHYRILEPROTHRSERPHCYSESLYS | 150 |
| 65 | TACTCCTATCAATTCGTCACAGNATNTCTCNAATTTTGTATACAAA | 450 |
| 66 | ASMSERTHGLYPHSSERTYRLEUVALILEARGLULEUGINALASE | 167 |
| 67 | AAATTCACACGGTTTTAGGTACCTGTCATAAGAAATAATTCGAACCTTC | 500 |
| 68 | IASPLEUALAGLYGLYCYSARGVALGLUTHVALAATRPSERSEIALAP | 184 |
| 69 | GGATTTGGCTGGCGGCTGAGGGTTGAACATGTTGCAATGGTCTCTGCTC | 550 |
| 70 | ROGLYLERSESRASNLYSESRTHRLEUSERSETHRHISGLINGLY | 200 |
| 71 | CAGGCAATTCATCAACACAAAGTCGTACGTTATCAACACACATCAAGGC | 600 |
| 72 | LEUALATYRGLYPHEGLULLEUSERTRPLYSARGASNLEUCYSARGAS | 217 |
| 73 | CTGGCTTATAGGGTTTGACCTTTCTGCAAGCGTATCTGTATGTGTAAGA | 650 |
| 74 | NCYSASPARYSERATRGVLYGLYGLYCYSTHILEGLINGLNASERTASPA | 234 |
| 75 | TTTCGACGCCGAATCTGTGGGGGTGAGTCGACATTTGAGNAAAACGGACCA | 700 |
| 76 | RGALATHRCYSARGTYRTPCYSLYSGILNASPILEHISVALSERLYLEU | 250 |
| 77 | GACCTACTTGCTGTTATGTGGTCAAGGACATTCACGTTTGGAAAGCTT | 750 |
| 78 | THPHEARGCYSLYVALGILTYTYSERVALYRVALLEUPHEHEGL | 267 |
| 79 | ACGTTCCGATCGCAAGTCGAGACATATCTGTATATGTATGTTCTTTGG | 800 |
| 80 | YGLYILEGLYILEGLYVALLEUALA**ARGPHEULEUGLYILEP | 284 |
| 81 | CGGTAAAGGAANAAGTGGAGTTTGGCCCTAAGATTTCTACTGTGGAATTC | 850 |
| 82 | ROLLEULEULEALALAVALTTPGILNCYSLYARGVALASNLNHS | 300 |
| 83 | CATCTTTGATGCCACAGTGGGTGGCGCAGTGCACAAAACGGAATTTGCAT | 900 |
| 84 | THNSERTASGLINGLNASNCYS | 308 |
| 85 | ACATCTCCGGAAGAACAGAAAGCT | 924 |

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:ABA01665

```

seq_documentation_block:
ID ABA01665 strand: DNA; 1210 BP.
AC ABA01665;
XX
XX 06-FEB-2002 (first entry)
XX
XX Nicotiana tabacum cv. Xanthine stress induced gene SEQ ID NO:3.
XX
XX Nicotiana tabacum cv. Xanthine; stress induced protein; plant;
XX stress induced gene; transgenic plant; resistance; ds.
XX
XX Nicotiana tabacum.
XX
XX KEY Location/Qualifiers
XX FT CDS 18..943
XX FT /*tag=a
XX FT /product="stress induced protein"
XX FT /trans_except= (pos:846..848,aa:Xaa)
XX
XX JP2001252084-A.
XX
XX 18-SEP-2001.
XX
XX 15-MAR-2000; 2000JP-0071655.
XX
XX 15-MAR-2000; 2000JP-0071655.
XX
XX (NARA-) NARA SENTAN KAKAKU GIUTSU DAIGAKUIN DAI.
XX
XX WPI; 2002-029660/04.
XX
XX P-PSDB; AAG68236.
XX
XX New polypeptide for creating plant resistance to injury, comprises a
XX PT gene, the expression of which is induced by stress
XX
XX Claim 4; Page 8; 11pp; Japanese.
XX
XX
XX The present sequence represents a gene isolated from tobacco, in which
XX CC the expression is induced by stress, and so encodes a stress induced
XX CC protein (1). The present invention also describes: (1) providing
XX CC resistance against stress by introducing the above gene to a plant; and
XX CC (2) a transformed plant which obtained resistance against stress by
XX CC introducing the above gene to the plant. The gene can be used for
XX CC creating plants resistant against injury, osmotic pressure, salt, or
XX CC low temperature-induced stress.
XX
XX Sequence 1210 BP; 347 A; 217 C; 251 G; 395 T; 0 other;
SQ
alignment_scores:
Quality: 1642.00 Length: 308
Ratio: 5.349 Gaps: 0
Percent Similarity: 99.675 Percent Identity: 99.675
alignment_block:
US-09-805-919-1 x ABA01665 ..
Align seg 1/1 to: ABA01665 from: 1 to: 1210
1 MetcteuThArGgLyLeuLeuPheAlaCyValLeuLeuValThLe 17
18 ATGTTGACAAAGAGGGCGCTTTCCCTTGCTGTTGTTGTTACTGTTGACACT 67
17 uIlSeSeSerLySaIaGlnAspIleSerGlnGlyValProSerSerC 34
68 CATTAAGCAATTCATTAAGCCGAGGATATTCATATGTCGCTTCTTCT 117
34 ySgIyAspIleGlnIleLySPhPeProPheArGLeuArGThAsPProGln 50
118 GCGGTGATTCATTAATTAATTTCCCTTCGCACTGGAGCATGATGCCGAG 167
51 HiscySgIyArGArGgLyTyRgIuLeuAspCySgIaNaenAnGlnThIva 67

```



```
|||||
168 CATTGGGTAGACGCGGATATGAGCTGATGGCCAGAACCAACCAACCGT 217
67 lPheasnTYrLYsSerArgIlePheaspValgInguileasnTYrArgS 84
218 GTCAATATCAATCCAGAAATTTTCAGCTACGAGAAATTAATCAAGAA 267
84 ertYSerIleargLeuleuLeuaspProglyLeuasnAspGlnArgGluasn 100
268 GCTACTCAATAGGCTACTTGACCGCGCTAAATGATCAGAGAAAT 317
101 CysThyValPheProasnHisArgAlaSerTYrAspAlaMetThrSerG 117
318 TGGCACTTTTCCAAATCAGAGCAACTATGATGCATACACACCA 367
117 nilePheGluTrpValArgValAsnAsnAspIleasnTYrValaAsnCYSL 134
368 AATCTTGAATGGGTTCGTGTACAAATGATATCAACTATGTCACACTGTC 417
134 euAlaProIleasnSerSerGlnTYrIleProThrSerPheCysSerLYs 150
418 TAGCTCTATCAATGCTACACTATATCTCAAGATTTTGTGCAAA 467
151 AsnSerThrGlyPheSerTYrLeuValIleArgGluIleLeuGlnAlaSe 167
468 AATCAACGCGGTTTACGTAAGTTCATACAGAAATATGCACAGCTTC 517
167 rAspLeuAlaGlyGlyCysArgValGluThrValAlaIatPserSerAlaP 184
518 GGATTTGGCTGGCGGCTGAGGGTTAAAGTGTGATGGTCTCTGCTGC 567
184 rocGlyIleSerSerAsnLYsSerSerThrLeuSerSerThrHisGlnGly 200
568 CAGGCAATTCATCAAAAGTGTCTACGTTATCAAGCACACATCAGGC 617
201 LeuAlaIArgGlyPheGluLeuSerTrpLYsArgAsnLeuLeuCYsArgAs 217
618 CTGGCTATGGGTTTGAAGCTTCTTGGAGCGCAATCTGTATGTAGAAA 667
217 nCYsAspArgSerArgGlyGlyGluCysThrIleGluGlnAsnSerAspA 234
668 TTGGAGACCGGAGTCGTGGGGGTAGTGCACATATGAGAGAAACAGCA 717
234 rGalaThrCysArgTYrTrpCysLYsGluAspIleHisValSerLYsLeu 250
718 GAGCTACTGTGCTTATGCTTGAAGAGACATTCACCTTCGAGAGCTT 767
251 ThrPheArgCysLYsValGluTYrTYrSerValTYrValLeuPhePheG 267
768 ACGTCCGATGCAAAAGTCAGTCTATCTCTGTTATGATGTTGTTCTTGG 817
267 yGlyIleGlyIleGlyGlyValLeuAla**ArgPheLeuLeuGlyIleP 284
818 CGGTATAGGAATAGTGGAGTGTGGCCCTAGAGATTTCTACACGAGATTC 867
284 roIleLeuIleAlaAlaValValTrpGlnCysLYsArgAsnLeuHis 300
868 CAATCTTGATGCGACAGAGTGGTGTGCAATGTCAGAAAGAGGAATTTGCAT 917
301 ThrSerSerAspGlnAsnCYs 308
918 ACATCTCGATGACAGAACTGT 941
seq_name: /stlsl/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AA54855
seq_documentation_block:
ID AA54855 standard; DNA; 1246 BP.
XX
AC AA54855;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 79315.
```

```
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
```

| PR | 23-JUN-1999 | 99US-0140353 | PR | 23-JUN-1999 | 99US-0140353 |
|----|-------------|--------------|----|-------------|--------------|
| PR | 23-JUN-1999 | 99US-0140353 | PR | 23-JUN-1999 | 99US-0140353 |
| PR | 24-JUN-1999 | 99US-0140695 | PR | 24-JUN-1999 | 99US-0140695 |
| PR | 26-JUN-1999 | 99US-0140823 | PR | 26-JUN-1999 | 99US-0140823 |
| PR | 30-JUN-1999 | 99US-0141287 | PR | 30-JUN-1999 | 99US-0141287 |
| PR | 01-JUL-1999 | 99US-0141842 | PR | 01-JUL-1999 | 99US-0141842 |
| PR | 02-JUL-1999 | 99US-0142154 | PR | 02-JUL-1999 | 99US-0142154 |
| PR | 06-JUL-1999 | 99US-0142390 | PR | 06-JUL-1999 | 99US-0142390 |
| PR | 08-JUL-1999 | 99US-0142803 | PR | 08-JUL-1999 | 99US-0142803 |
| PR | 09-JUL-1999 | 99US-0142820 | PR | 09-JUL-1999 | 99US-0142820 |
| PR | 12-JUL-1999 | 99US-0142877 | PR | 12-JUL-1999 | 99US-0142877 |
| PR | 13-JUL-1999 | 99US-0143542 | PR | 13-JUL-1999 | 99US-0143542 |
| PR | 14-JUL-1999 | 99US-0143624 | PR | 14-JUL-1999 | 99US-0143624 |
| PR | 15-JUL-1999 | 99US-0144005 | PR | 15-JUL-1999 | 99US-0144005 |
| PR | 16-JUL-1999 | 99US-0144085 | PR | 16-JUL-1999 | 99US-0144085 |
| PR | 19-JUL-1999 | 99US-0144325 | PR | 19-JUL-1999 | 99US-0144325 |
| PR | 19-JUL-1999 | 99US-0144331 | PR | 19-JUL-1999 | 99US-0144331 |
| PR | 19-JUL-1999 | 99US-0144332 | PR | 19-JUL-1999 | 99US-0144332 |
| PR | 19-JUL-1999 | 99US-0144334 | PR | 19-JUL-1999 | 99US-0144334 |
| PR | 19-JUL-1999 | 99US-0144334 | PR | 19-JUL-1999 | 99US-0144334 |
| PR | 19-JUL-1999 | 99US-0144335 | PR | 19-JUL-1999 | 99US-0144335 |
| PR | 20-JUL-1999 | 99US-0144352 | PR | 20-JUL-1999 | 99US-0144352 |
| PR | 20-JUL-1999 | 99US-0144632 | PR | 20-JUL-1999 | 99US-0144632 |
| PR | 21-JUL-1999 | 99US-0144814 | PR | 21-JUL-1999 | 99US-0144814 |
| PR | 21-JUL-1999 | 99US-0145086 | PR | 21-JUL-1999 | 99US-0145086 |
| PR | 21-JUL-1999 | 99US-0145088 | PR | 21-JUL-1999 | 99US-0145088 |
| PR | 22-JUL-1999 | 99US-0145085 | PR | 22-JUL-1999 | 99US-0145085 |
| PR | 22-JUL-1999 | 99US-0145087 | PR | 22-JUL-1999 | 99US-0145087 |
| PR | 22-JUL-1999 | 99US-0145089 | PR | 22-JUL-1999 | 99US-0145089 |
| PR | 22-JUL-1999 | 99US-0145192 | PR | 22-JUL-1999 | 99US-0145192 |
| PR | 23-JUL-1999 | 99US-0145218 | PR | 23-JUL-1999 | 99US-0145218 |
| PR | 23-JUL-1999 | 99US-0145224 | PR | 23-JUL-1999 | 99US-0145224 |
| PR | 26-JUL-1999 | 99US-0145276 | PR | 26-JUL-1999 | 99US-0145276 |
| PR | 27-JUL-1999 | 99US-0145913 | PR | 27-JUL-1999 | 99US-0145913 |
| PR | 27-JUL-1999 | 99US-0145918 | PR | 27-JUL-1999 | 99US-0145918 |
| PR | 28-JUL-1999 | 99US-0145951 | PR | 28-JUL-1999 | 99US-0145951 |
| PR | 02-AUG-1999 | 99US-0146386 | PR | 02-AUG-1999 | 99US-0146386 |
| PR | 02-AUG-1999 | 99US-0146388 | PR | 02-AUG-1999 | 99US-0146388 |
| PR | 03-AUG-1999 | 99US-0146389 | PR | 03-AUG-1999 | 99US-0146389 |
| PR | 04-AUG-1999 | 99US-0147038 | PR | 04-AUG-1999 | 99US-0147038 |
| PR | 04-AUG-1999 | 99US-0147204 | PR | 04-AUG-1999 | 99US-0147204 |
| PR | 05-AUG-1999 | 99US-0147302 | PR | 05-AUG-1999 | 99US-0147302 |
| PR | 05-AUG-1999 | 99US-0147260 | PR | 05-AUG-1999 | 99US-0147260 |
| PR | 06-AUG-1999 | 99US-0147303 | PR | 06-AUG-1999 | 99US-0147303 |
| PR | 06-AUG-1999 | 99US-0147416 | PR | 06-AUG-1999 | 99US-0147416 |
| PR | 09-AUG-1999 | 99US-0147493 | PR | 09-AUG-1999 | 99US-0147493 |
| PR | 09-AUG-1999 | 99US-0147935 | PR | 09-AUG-1999 | 99US-0147935 |
| PR | 10-AUG-1999 | 99US-0148171 | PR | 10-AUG-1999 | 99US-0148171 |
| PR | 11-AUG-1999 | 99US-0148319 | PR | 11-AUG-1999 | 99US-0148319 |
| PR | 12-AUG-1999 | 99US-0148341 | PR | 12-AUG-1999 | 99US-0148341 |
| PR | 13-AUG-1999 | 99US-0148565 | PR | 13-AUG-1999 | 99US-0148565 |
| PR | 13-AUG-1999 | 99US-0148684 | PR | 13-AUG-1999 | 99US-0148684 |
| PR | 16-AUG-1999 | 99US-0149368 | PR | 16-AUG-1999 | 99US-0149 |

72 erArgIlePheAspValGlnGluIleAsnTyrArgSerTyrSerIleArg 88
||||| ||| ||||| |||||
970 CCAGACGATTTCTCTCCACACATTA.....AAGTCGACGCGTATTCAC 927
89 LeuLeuAsnProGlyLeuAsnAspGlnArgGluAsnGlyThrValPhePr 105
||||| ||||| |||||
926 CTTAGAGACCCC.....GATTAAGTGC.....TTAGC 901
105 oAsnHisArgAlaSerTyrAspAlaMetThrSerGlnIlePheGluTyrV 122
||||| ||||| |||||
900 CAGACGCGCTTCGACGCTGACGCTCAATGCTCTCTCTCTCTCTCTCTC 851
122 aIArgValAsnAsnAspIleAsnTyrValAsnGlyLeuAlaIleAsn 138
||||| ||||| |||||
850 ACCTTGTCAC.....TACACTTCTTGAGCTGCTCATTAACGAAACGTC 807
139 SerSerGlnTyrIleProThrSerPheCysSerTyrAsnSerThrGlyPh 155
||||| ||||| |||||
806 AATTCCTCTCTCTTGGAACCCATCCATGCTCTTACTACTATTCACA..... 762
155 eSerTyrLeuValIleArgGluIleLeuGlnAlaSerAspLeuAlaGlyG 172
||||| ||||| |||||
761CCTTCAGATCTCACCGGTT 743
172 Ly.CysArg.....ValGluThrVal 178
||||| ||||| |||||
742 CAATGCCGCTCTCTGTGCACACCTTGACACACTACTTCTCCGGCTCT 693
178 IAlaTyrSerSerAlaProGlyIleSerSerAsnLysSerSerThrLeuS 195
||||| ||||| |||||
692 TCACCCATCCGACGCTG.....ATCTCAACAACACAGACCTCTGCTGAA 649
195 eSerThrHisGln 199
||||| ||||| |||||
648 ATGGGACTCACCGG 635
seq_name: /sids1/gcdata/hold-geneseq/geneseqn-emb1/NM2000.DAT:AAC52806
seq_documentation_block:
ID AAC52806 strand: DNA; 1249 bp.
NC AAC52806;
XX
XX
DT 18-OCT-2000 (first entry)
XX
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 72505.
XX
XX KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
XX PN EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0125264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130445.
PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144088.
PR 16-JUL-1999; 99US-0144088.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.


```

:: |||||
744 CAATGCCGCTCTGTCTCAATCTCAAGACCGCTACTCTCTCCAGCTTAT 695
178 lalatpserSerAlaProGlyIleSerSerAnlySerSerThrLeus 195
    |||||
694 TCACCACTCGCAGTTG.....ACCTCAACAACAAGACCTCTGCTCTAA 651
195 erSerThrHsGln 199
    ::|||
650 ATGGAGCTCACCAG 637

seq_name: /SID1/gcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AC53268
seq_documentation_block:
ID AC53268 standard; DNA; 1177 BP.
XX
AC AC53268;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 73938.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132853.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
```

```

PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
```

| | | |
|----|--------------|---------------|
| PR | 09-AUG-1999; | 99US-0147493. |
| PR | 09-AUG-1999; | 99US-0147935. |
| PR | 10-AUG-1999; | 99US-0148171. |
| PR | 11-AUG-1999; | 99US-0148319. |
| PR | 12-AUG-1999; | 99US-0148341. |
| PR | 13-AUG-1999; | 99US-0148565. |
| PR | 13-AUG-1999; | 99US-0148684. |
| PR | 16-AUG-1999; | 99US-0149368. |
| PR | 17-AUG-1999; | 99US-0149375. |
| PR | 18-AUG-1999; | 99US-0149426. |
| PR | 20-AUG-1999; | 99US-0149722. |
| PR | 20-AUG-1999; | 99US-0149723. |
| PR | 20-AUG-1999; | 99US-0149929. |
| PR | 23-AUG-1999; | 99US-0149929. |
| PR | 23-AUG-1999; | 99US-0149930. |
| PR | 25-AUG-1999; | 99US-0150566. |
| PR | 26-AUG-1999; | 99US-0150884. |
| PR | 27-AUG-1999; | 99US-0151065. |
| PR | 27-AUG-1999; | 99US-0151066. |
| PR | 27-AUG-1999; | 99US-0151080. |
| PR | 30-AUG-1999; | 99US-0151303. |
| PR | 31-AUG-1999; | 99US-0151438. |
| PR | 01-SEP-1999; | 99US-0151930. |
| PR | 07-SEP-1999; | 99US-0152363. |
| PR | 10-SEP-1999; | 99US-0153070. |
| PR | 13-SEP-1999; | 99US-0153758. |
| PR | 15-SEP-1999; | 99US-0154018. |
| PR | 16-SEP-1999; | 99US-0154039. |
| PR | 20-SEP-1999; | 99US-0154779. |
| PR | 22-SEP-1999; | 99US-0155139. |
| PR | 23-SEP-1999; | 99US-0155486. |
| PR | 24-SEP-1999; | 99US-0155659. |
| PR | 28-SEP-1999; | 99US-0156458. |
| PR | 29-SEP-1999; | 99US-0156599. |
| PR | 04-OCT-1999; | 99US-0157117. |
| PR | 05-OCT-1999; | 99US-0157753. |
| PR | 06-OCT-1999; | 99US-0157863. |
| PR | 07-OCT-1999; | 99US-0158029. |
| PR | 08-OCT-1999; | 99US-0158232. |
| PR | 12-OCT-1999; | 99US-0158369. |
| PR | 13-OCT-1999; | 99US-0159293. |
| PR | 13-OCT-1999; | 99US-0159294. |
| PR | 13-OCT-1999; | 99US-0159295. |
| PR | 14-OCT-1999; | 99US-0159329. |
| PR | 14-OCT-1999; | 99US-0159330. |
| PR | 14-OCT-1999; | 99US-0159331. |
| PR | 14-OCT-1999; | 99US-0159637. |
| PR | 18-OCT-1999; | 99US-0159538. |
| PR | 21-OCT-1999; | 99US-0159584. |
| PR | 21-OCT-1999; | 99US-0160741. |
| PR | 21-OCT-1999; | 99US-0160767. |
| PR | 21-OCT-1999; | 99US-0160768. |
| PR | 21-OCT-1999; | 99US-0160770. |
| PR | 21-OCT-1999; | 99US-0160814. |
| PR | 21-OCT-1999; | 99US-0160815. |
| PR | 22-OCT-1999; | 99US-0160980. |
| PR | 22-OCT-1999; | 99US-0160981. |
| PR | 22-OCT-1999; | 99US-0160989. |
| PR | 25-OCT-1999; | 99US-0161404. |
| PR | 25-OCT-1999; | 99US-0161405. |
| PR | 25-OCT-1999; | 99US-0161406. |
| PR | 26-OCT-1999; | 99US-0161359. |
| PR | 26-OCT-1999; | 99US-0161360. |
| PR | 26-OCT-1999; | 99US-0161361. |
| PR | 28-OCT-1999; | 99US-0161920. |
| PR | 28-OCT-1999; | 99US-0161922. |
| PR | 28-OCT-1999; | 99US-0161993. |
| PR | 29-OCT-1999; | 99US-0162142. |

```

alignment_block:
US-09-805-919-1 x AAC53268      ..
Align seg 1/1 to: AAC53268 from: 1 to: 1177

```

6 Leu⁶Leu⁷Val⁸Leu⁹Leu¹⁰Val¹¹Thr¹²Leu¹³Ser¹⁴Ser¹⁵Leu¹⁶Val¹⁷Leu¹⁸Val¹⁹Val²⁰Val²¹Val²²Val²³Val²⁴Val²⁵Val²⁶Val²⁷Val²⁸Val²⁹Val³⁰Val³¹Val³²Val³³Val³⁴Val³⁵Val³⁶Val³⁷Val³⁸Val³⁹Val⁴⁰Val⁴¹Val⁴²Val⁴³Val⁴⁴Val⁴⁵Val⁴⁶Val⁴⁷Val⁴⁸Val⁴⁹Val⁵⁰Val⁵¹Val⁵²Val⁵³Val⁵⁴Val⁵⁵Val⁵⁶Val⁵⁷Val⁵⁸Val⁵⁹Val⁶⁰Val⁶¹Val⁶²Val⁶³Val⁶⁴Val⁶⁵Val⁶⁶Val⁶⁷Val⁶⁸Val⁶⁹Val⁷⁰Val⁷¹Val⁷²Val⁷³Val⁷⁴Val⁷⁵Val⁷⁶Val⁷⁷Val⁷⁸Val⁷⁹Val⁸⁰Val⁸¹Val⁸²Val⁸³Val⁸⁴Val⁸⁵Val⁸⁶Val⁸⁷Val⁸⁸Val⁸⁹Val⁹⁰Val⁹¹Val⁹²Val⁹³Val⁹⁴Val⁹⁵Val⁹⁶Val⁹⁷Val⁹⁸Val⁹⁹Val¹⁰⁰Val¹⁰¹Val¹⁰²Val¹⁰³Val¹⁰⁴Val¹⁰⁵Val¹⁰⁶Val¹⁰⁷Val¹⁰⁸Val¹⁰⁹Val¹¹⁰Val¹¹¹Val¹¹²Val¹¹³Val¹¹⁴Val¹¹⁵Val¹¹⁶Val¹¹⁷Val¹¹⁸Val¹¹⁹Val¹²⁰Val¹²¹Val¹²²Val¹²³Val¹²⁴Val¹²⁵Val¹²⁶Val¹²⁷Val¹²⁸Val¹²⁹Val¹³⁰Val¹³¹Val¹³²Val¹³³Val¹³⁴Val¹³⁵Val¹³⁶Val¹³⁷Val¹³⁸Val¹³⁹Val¹⁴⁰Val¹⁴¹Val¹⁴²Val¹⁴³Val¹⁴⁴Val¹⁴⁵Val¹⁴⁶Val¹⁴⁷Val¹⁴⁸Val¹⁴⁹Val¹⁵⁰Val¹⁵¹Val¹⁵²Val¹⁵³Val¹⁵⁴Val¹⁵⁵Val¹⁵⁶Val¹⁵⁷Val¹⁵⁸Val¹⁵⁹Val¹⁶⁰Val¹⁶¹Val¹⁶²Val¹⁶³Val¹⁶⁴Val¹⁶⁵Val¹⁶⁶Val¹⁶⁷Val¹⁶⁸Val¹⁶⁹Val¹⁷⁰Val¹⁷¹Val¹⁷²Val¹⁷³Val¹⁷⁴Val¹⁷⁵Val¹⁷⁶Val¹⁷⁷Val¹⁷⁸Val¹⁷⁹Val¹⁸⁰Val¹⁸¹Val¹⁸²Val¹⁸³Val¹⁸⁴Val¹⁸⁵Val¹⁸⁶Val¹⁸⁷Val¹⁸⁸Val¹⁸⁹Val¹⁹⁰Val¹⁹¹Val¹⁹²Val¹⁹³Val¹⁹⁴Val¹⁹⁵Val¹⁹⁶Val¹⁹⁷Val¹⁹⁸Val¹⁹⁹Val²⁰⁰Val²⁰¹Val²⁰²Val²⁰³Val²⁰⁴Val²⁰⁵Val²⁰⁶Val²⁰⁷Val²⁰⁸Val²⁰⁹Val²¹⁰Val²¹¹Val²¹²Val²¹³Val²¹⁴Val²¹⁵Val²¹⁶Val²¹⁷Val²¹⁸Val²¹⁹Val²²⁰Val²²¹Val²²²Val²²³Val²²⁴Val²²⁵Val²²⁶Val²²⁷Val²²⁸Val²²⁹Val²³⁰Val²³¹Val²³²Val²³³Val²³⁴Val²³⁵Val²³⁶Val²³⁷Val²³⁸Val²³⁹Val²⁴⁰Val²⁴¹Val²⁴²Val²⁴³Val²⁴⁴Val²⁴⁵Val²⁴⁶Val²⁴⁷Val²⁴⁸Val²⁴⁹Val²⁵⁰Val²⁵¹Val²⁵²Val²⁵³Val²⁵⁴Val²⁵⁵Val²⁵⁶Val²⁵⁷Val²⁵⁸Val²⁵⁹Val²⁶⁰Val²⁶¹Val²⁶²Val²⁶³Val²⁶⁴Val²⁶⁵Val²⁶⁶Val²⁶⁷Val²⁶⁸Val²⁶⁹Val²⁷⁰Val²⁷¹Val²⁷²Val²⁷³Val²⁷⁴Val²⁷⁵Val²⁷⁶Val²⁷⁷Val²⁷⁸Val²⁷⁹Val²⁸⁰Val²⁸¹Val²⁸²Val²⁸³Val²⁸⁴Val²⁸⁵Val²⁸⁶Val²⁸⁷Val²⁸⁸Val²⁸⁹Val²⁹⁰Val²⁹¹Val²⁹²Val²⁹³Val²⁹⁴Val²⁹⁵Val²⁹⁶Val²⁹⁷Val²⁹⁸Val²⁹⁹Val³⁰⁰Val³⁰¹Val³⁰²Val³⁰³Val³⁰⁴Val³⁰⁵Val³⁰⁶Val³⁰⁷Val³⁰⁸Val³⁰⁹Val³¹⁰Val³¹¹Val³¹²Val³¹³Val³¹⁴Val³¹⁵Val³¹⁶Val³¹⁷Val³¹⁸Val³¹⁹Val³²⁰Val³²¹Val³²²Val³²³Val³²⁴Val³²⁵Val³²⁶Val³²⁷Val³²⁸Val³²⁹Val³³⁰Val³³¹Val³³²Val³³³Val³³⁴Val³³⁵Val³³⁶Val³³⁷Val³³⁸Val³³⁹Val³⁴⁰Val³⁴¹Val³⁴²Val³⁴³Val³⁴⁴Val³⁴⁵Val³⁴⁶Val³⁴⁷Val³⁴⁸Val³⁴⁹Val³⁵⁰Val³⁵¹Val³⁵²Val³⁵³Val³⁵⁴Val³⁵⁵Val³⁵⁶Val³⁵⁷Val³⁵⁸Val³⁵⁹Val³⁶⁰Val³⁶¹Val³⁶²Val³⁶³Val³⁶⁴Val³⁶⁵Val³⁶⁶Val³⁶⁷Val³⁶⁸Val³⁶⁹Val³⁷⁰Val³⁷¹Val³⁷²Val³⁷³Val³⁷⁴Val³⁷⁵Val³⁷⁶Val³⁷⁷Val³⁷⁸Val³⁷⁹Val³⁸⁰Val³⁸¹Val³⁸²Val³⁸³Val³⁸⁴Val³⁸⁵Val³⁸⁶Val³⁸⁷Val³⁸⁸Val³⁸⁹Val³⁹⁰Val³⁹¹Val³⁹²Val³⁹³Val³⁹⁴Val³⁹⁵Val³⁹⁶Val³⁹⁷Val³⁹⁸Val³⁹⁹Val⁴⁰⁰Val⁴⁰¹Val⁴⁰²Val⁴⁰³Val⁴⁰⁴Val⁴⁰⁵Val⁴⁰⁶Val⁴⁰⁷Val⁴⁰⁸Val⁴⁰⁹Val⁴¹⁰Val⁴¹¹Val⁴¹²Val⁴¹³Val⁴¹⁴Val⁴¹⁵Val⁴¹⁶Val⁴¹⁷Val⁴¹⁸Val⁴¹⁹Val⁴²⁰Val⁴²¹Val⁴²²Val⁴²³Val

257 lufTyTyrSerValTyrValLeuPhePheGlyIleGlyIleGlyIle 273
 : : : : :
 873 GATGGATA.....ATAGTGGT 889
 274 ValLeuAla***ArpPheLeuLeuGlyIleProIleLeuIleAlaVala 290
 |||||
 890 GTGTGATC.....TTGATCTCATAGTATCAGGTATATTGTTCCT 933
 290 ValTTPGlnCys 294
 : : : : :
 934 GGTGTTCTTTC 946

seq_name: /sids1/gcgdata/hold-genesec/geneseqn-emb1/NA2001B.DAT:ABL16690

seq_documentation_block:
 ID ABL16690 standard; DNA; 4169 BP.

XX ABL16690;
 AC
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 1543.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 PE 23-MAR-2001; 2001MO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1: SEQ ID NO 1543; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABJ01840-ABJ16175) and the encoded proteins
 CC (ABJ5737-ABJ72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_prt_sequences.
 CC
 XX
 SQ Sequence 4169 BP; 1246 A; 859 C; 919 G; 1143 T; 0 other;

alignment_scores:

Quality: 105.50 Length: 301
 Ratio: 0.805 Gaps: 15
 Percent Similarity: 43.522 Percent Identity: 20.930

alignment_block:
 US-09-805-919-1 x ABL16690/rev ..

Align seg 1/1 to reverse of: ABL16690 from: 1 to: 4169

19 SerSerSerIysAlaGln.....AspIleSerGI 28
 |||:|||||
 3176 TCACACCTTAAGTACCAACCTAAATATTTGCTGTGGCAATCTGA 3127
 28 nCysValProSerSerCysGlyAspIleGlnIleIysPheProPheArgL 45
 :|||:|||||
 3126 GTGTTGGCCACAAACTGTTGAAAAACGAAATCCAG..... 3090
 45 euArgThrAspProGlnHisCysGlyArg. ArgGlyTyrGlnLeuAspCys 61
 ||| :|||:|||||
 3089 ..CGAGGCAACCGCAATCAGGCCGCGACAGCGAGTCAACAGAAAGCGC 3042
 61 sGlnAsn.....AsnG 65
 :|||:|||||
 3041 GCAAACGAGTCCACACCTGCACAAACATGGGACGAGCGCCAGCGC 2992
 65 lnrValPheAsnTyrIysSerArgIlePheAspValGlnGlnLeuAsn 81
 |||:|||||
 2991 AAACCAAG.....GTGCCG 2978
 82 TyrArgSerTyrSerIleArgLeuLeu. AspProGly.....LeuAsnA 96
 :|||:|||||
 2977 TTCGCTGCAATTTGTTATGCTGTGATGCGCCAGCTGTACGTGCTCCT 2928
 96 spGlnArgGlnAsnCysThrValPheProAsnHisArgAlaSerTyrTrp 112
 :|||:|||||
 2927 CTCCCAAGCAGATGAGCACCAATCACAGGTGATGCCATGCCATATCC 2878
 113 AlameThrSerGlnIlePheGlnTyrValArgValAsnAsp..... 127
 :|||:|||||
 2877 TCAGACACCGAATACATGATCCACTTGCACGAGTACATAGACGGGA 2828
 128IleAsnTyrValAsnCysLeuAlaProIleAsnSerSerGlnT 142
 :|||:|||||
 2827 GGAGGTGTGATCTGTGATGATGACACCGGTGGCCGATACCAATCGCGAG. 2779
 142 YrIleProThrSerPheCysSerIysAsnSerThrGlyPheSerTyrLeu 158
 2779 2779
 159 ValIleArgGlnIleLeuGlnIleAspLeuAlaGlyIleCysArgVala 175
 2779 2779
 175 lGlnThrValAlaIrpSerSerAlaProGlyIleSerSerAsnIysers 192
 :|||:|||||
 2778 .GAGACGTACGCTACTTCTCTCTCCCTTTTGCAGTGGCCAGAGTCTCT 2730
 192 er.....ThrLeuSerSerThrHisGlnIleLeuAla 202
 :|||:|||||
 2729 CGATATGCACTACCAACGAGACCTGTAGGAGCGCTGCMAAGAGTCTGAG 2680
 203 Tyr.....GlyPheGlnLeuSerTrpIys.....ArgAs 212
 :|||:|||||
 2679 CTAGAGTTCAGTGTGCTACGATGAGTTCAGAGGAGCGAGCGCCCAAAATC 2630
 212 nLeuLeuCysArgAsnCysAspArgSerArgIleGlyIleGlyIleGlyIle 229
 :|||:|||||
 2629 GGCATCTGCATG.....GTACACTGCG 2607
 229 lueGlnAsnSerAspArgAlaThrCysArgTyrTrpCysIysLeuAspIle 245
 :|||:|||||
 2606 AGGAGGAGAGCGCCAAAGCA..... 2587
 246 HisValSerIysLeuThrPheArgCysIysValGlnTyrTyrSerValTy 262
 :|||:|||||
 2586TTCACTATGCCGCGAAGAAAGAGATACTGG.....TA 2555
 262 rValLeuPhePheGlyIleGlyIleGlyIleGlyIleValLeuAla***Arg 278
 :|||:|||||
 2554 CCAGATGTATCATCATGATGAGACTGCCATTGGGAAAAAGTGGTGAAGCGC 2506


```
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152362.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161989.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
alignment_scores:
  Quality: 104.00      Length: 110
  Ratio: 1.651        Gaps: 7
Percent Similarity: 57.273  Percent Identity: 31.818
```

alignment_block:

US-09-805-919-1 x AAC54200/rev ..

Align seg 1/1 to reverse of: AAC54200 from: 1 to: 494

```
1 MetLeuHrArgGlyLeuLeuPheAlaCysValLeuLeuValThrIle 17
  :::::::::::::::::::::
373 TTATATCATCTCTCTCTGTCTTCTATCTCTATATCCGACGACGAC 324
```

```
17 uileSerSerLysAlaGlnAspIleSerGlnCysValProSerC 34
  :::::::::::
323 GAGCACCACCTCCACCGTT.....ACATGCACCTAACGCCGTTT 286
  :::::::::::
34 ys.....GlyAspIleGlnIleLysPheProPheArgThr 47
  :::::::::::
285 GTGCCCGGTACGCCCGGATC...ATCCGTTTCCATCCGCTATAAAC 239
  :::::::::::
48 AspProGlnHis...CysGly...ArgArgGlyTyrGlnLeuAspCysG 62
  :::::::::::
238 CAACAATCATCTCTCTGTCTTACGCAAAAGCGCTCAAGCTTAACGTGCG 189
  :::::::::::
62 n.....AsnAsnGlnThrValPheAsnTyrIysSerArgIleP 75
  :::::::::::
188 TATTAAGCCGGTACAGACACCATACGTTACCGTTTCCGCTAACT 139
  :::::::::::
75 heAspValGlnGlnIleAsnTyrArgSerTyrSerIleArgLeuAsp 91
  :::::::::::
138 TCACCGTGCAGAGATCGATTCAGCGACGTCACAGATTTGATCAGCAC 89
  :::::::::::
92 ProGlyLeuAsnAspGlnArgGlnAsnCys 101
  :::::::::::
88 CCA.....AACAACTGT 77
  :::::::::::
seq_name: /sids1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC52497
seq_documentation_block:
ID AAC52497 standard; DNA; 497 BP.
AC AAC52497;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 71540.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
```

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135333.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141289.
PR 01-JUL-1999; 99US-0141844.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143544.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147039.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147937.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155119.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.


```
5237 ACTGCTCTTACCTGCCGTAAAGCAGGTGAGACTCAGCGGACCAACCTCAATT 5286
159 .....ValIleArgIu.....IleuGlnA 166
5287 AGAAGTACGACGCGTGGAGGGCTTTCGAGAGAGAGGCTATTGCAATT 5336
166 IAserAspleuAlagly.....GlyCysArgValGluThrValAlaIrrp 180
5337 CGTGGGGTTCCCGGCACTACCTGGCTGGCTTGTGCGAGTGGGTGGCCACC 5386
181 SerSerAlProGlyIleSerSerAnLysSerSerThrLeuSerSerThr 197
5387 ACATCAATGCCCGTGGTGAGAGACTCCACATCTCATATTATTTCCGCCAC 5436
197 rHisGlyIleuAlaIrrpGlyPheGlyLeuSer..... 208
5437 TCCCTGGCGGACAGCGTTGAGCTTCAACTCCAGCTGAGAGTTCCTCTGT 5486
209 .....TrpLysArgAsnLeuLeuCysArgAsnCys 218
5487 GCAGCGGACCAATCGATTGACATCGTTGAGATGCTGTGACGCTCATCG 5536
219 AspArgSerArgIleGlyIleGlyCysTrpIleGlu..... 229
5537 CAGCGGCGCAGGCGCATCGCGCTCATCTCATCGCGCTGCGCGCGCTC 5586
230 .....GluAsnSerAspArgAlaThrCysArgTrpIrrpCysL 242
5587 TGATTGCTCGGCATCGAATCGACGCTGCGACGCGCATGCGCATGCTGCA 5636
242 ysgLusAspIleHisValSerLysLeuThrPheArgCysLysValGluTrp 258
5637 GCAC.....CACGCGGCGGCGGCGCTCGCTCGCTCGCGCCACACGC 5679
259 TyrSerValIrrpValLeuPhePheGlyIleGlyIleGlyIleGlyValle 275
5680 TACATCGAATGGCGCAGCTCATCGGTAGTT.....GGCAACAC 5720
275 uAla**ArgPheLeuLeuGlyIleProIleLeuIleAla.....Alav 290
5721 GCGGTTCATCTTCATCATCGTTAATGTTGCTTCCTCATCTGATTGGCA 5770
290 aIValIrrpGlyCysLysArg 296
5771 TCACGTGACAGCTGTGCGCGG 5790
seq_name: /stids/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC39355
seq_documentation_block:
ID AAC39355 standard; DNA; 1166 bp.
XX
AAC39355;
XX
17-OCT-2000 (first entry)
XX
Arabidopsis thaliana DNA fragment SEQ ID NO: 24330.
DE
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
PN EP103405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
```

```
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139499.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140033.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141844.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
```



```

149 eRlysaNsErThrGly.....PheSerTyrlEuValIleArgInu 162
546 ATGCTCAACCTCTCTCCATTGACCTTCTTCTACTCAAAATTCAGGGCC 595
163 IleuGlnAlaSerAspLeuAlaGlyGlyCysArgValGluThrValAl 179
596 ATAGCCTCATTAACCTCCAGATACAGCTGTGGTC...TAGCTGCC 642
179 aTPSeSerAlaProGlyIleSerSerAsnLysSerSerThrLeuSer 196
643 TCTAGACCTCGACCATCGTTCGAATGATTGAAACAACCAAGTGTCT 692
196 eRThrsInGlyLeu..... 201
693 CCGTATACCTCGGCTTCTCAACCTCGGTCCAGGCCCAAGATCCACCG 742
202 .....AlaTyrlGlyPheGluLeuSerTrpLysArgAsnLeu.... 213
743 GAGAAATTGAACACTACGGAATCGCATCAAGTCAAACTCAACGTTTGA 792
214 .....LeuCySArGAsnCysAspArgSerArgGlyGlyGluC 226
793 TGAATTACCTCGGCTCTCGGCTAGCTCCAGCTTACTTAAGGTCT...T 839
226 yStrIleGluGluAsnSerAspArgAlaThrCysArgTyTrpCysLys 242
840 GTGGCTTCAACACACACACTCAAGCTCGTGTGTATAC.....TGTCCT 883
243 GluAspIleHisValSerLysLeuThrPhe.....Ar 253
884 GTGGGATTAACACGACATCGATGCTCTCTCTTCTTAATCTCGCTTC 933
253 gCyLysValGluTyTrSer.....ValTyrlValLeuPheP 266
934 TATCTGTCTCTCTGTCTCAAGAGATGGTATATACCTTGTCTAT 983
266 heGlyGlyIleGlyIleGlyValLeuAla**ArgPheLeuGly 282
984 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1033
283 IleProIleLeuIleAlaValAlaValIleGluCysLysArgArgAsn 299
1034 .....TGAAGAGTTCAAAGAGAGAGCT 1056
299 uHisThrSerSer 303
1057 TAATGATCGAAT 1069

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL16691
seq_documentation_block:
ID ABL16691 standard; DNA; 1863 BP.
XX
XX ABL16691;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 1546.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX
XX pharmacological; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX W0200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.

```

```

XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW,
XX
XX WPI, 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 1546; 21pp + sequence listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (AB57737-AB572072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 1863 BP; 419 A; 496 C; 486 G; 462 T; 0 other;

alignment_scores:
Quality: 99.50 Length: 268
Ratio: 0.816 Gaps: 13
Percent Similarity: 45.522 Percent Identity: 21.642

alignment_block:
US-09-805-919-1 x ABL16691 ...

Align seg 1/1 to: ABL16691 from: 1 to: 1863

27 SerGlnCysValProSerSerCysGlyAspIleGlnIleLysPheProph 43
|||||:|||||:|||||:|||||:|||||:
39 TCTGAGCTGTGGCCACAAACTGTTGAAAACCAATCCCAAC..... 80
43 eArgLeuArgThrAspProGluHisCysGlyArg ArgGlyTyrlGluLeu 59
|||||:|||||:|||||:|||||:|||||:
81 .....CGAGGCCAACCGATCAGCGCGCAGAGGAGGACCAAGAA 123
60 AspCysGlnAsnAsnGlnThrValPheAsnTyLysSerArgIlePheAs 76
|||||:|||||:|||||:|||||:
124 GACCGCCCAACGAGG..... 138
76 pValGlnGluIleAsnTyArgSerTySerIleArgLeuLeuAspPro 93
|||||:|||||:|||||:|||||:
139 .....TACCAC 145
93 LysLeuAsnAspGlnArgGluAsnCysThrValPheProAsnHisArgAla 109
|||||:|||||:|||||:|||||:
146 CTGCACAAGACACA.....ATGGGC 165
110 SerTyrlAspAlaMetThrSerGlnIlePheGluTrpValArgValAsnAs 126
|||||:|||||:|||||:|||||:
166 AGCAGCCGCACGCAACACGAGTGCCTCCGTCGCAATT..... 204
126 nasPileAsnTyValAsnCysLeuAlaProIleAsnSerSerGlnTyrl 143
|||||:|||||:|||||:|||||:
205 .....TGTTATGCTCTGCTGATCGCAGCTGCTAG 235
143 leProThrSerPheCysSerLysAsnSerThrGlyPheSerTyrlEuVal 159
|||||:|||||:|||||:|||||:
236 TGTGCTCTCCAGGACGATGAGCACAAAT.....CACAACTACATGAC 279
160 IleArgGluIleLeuGlnAlaSerAspLeuAlaGlyGlyCys.....Ar 174
|||||:|||||:|||||:|||||:
280 CGGAGAGAGGTGTACTGTGATGAACAGCGTGGCCCGTACCACAAATG 329

```


173 ysarYvalgluThrValAlaTTPSerSerAlaProGlyIleSerSerAsn 189
 563
 190 LysSerSerThrLeuSerSerThrHisGlnGlyLeuAlaTyrGlyPheG1 206
 581 AGCAGCGGAGCATCGCTTCTCCGAGAGATGATTAGATGATGCGCGA 630
 206 u.....LeuSerTTPLeuArgAsnLeuLeuGlyArgAsnGlySapArgS 221
 631 GGTAAATTGCTGCTCGGAAAGCCG...GACTGCAAGCTATGACAGAGA 677
 221 erarGgIyGlyGlyCysThrIleGluGluAsnSerAspArg 234
 678 AC...GGTGTGTGTGTGCGGGGAGAGAGACGATCTCCG 715

seq_name: /SIDSI/gcdata/hold-geneseq/gene-seq-emb1/NA2001B.DAT: AAS92969

seq_documentation_block:

ID AAS92969 standard; cDNA; 3874 BP.

AC AAS92969;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #28773.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR P-PSDB; ABG28782.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 1; SEQ ID No 28773; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human

CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 3874 BP; 824 A; 1213 C; 1059 G; 778 T; 0 other;

alignment_scores:

| Quality: | 95.00 | Length: | 341 |
|---------------------|--------|-------------------|--------|
| Ratio: | 0.605 | Gaps: | 20 |
| Percent Similarity: | 46.041 | Percent Identity: | 20.821 |

alignment_block:

US-09-805-919-1 x AAS92969 ..

Align seg 1/1 to: AAS92969 from: 1 to: 3874

29 CysValPro.....SerSerCysGlyAspIleGlnIleLysPhe 41

2495 TGGCTTCCTCGGGAAGGCGCTGCGAGCGCTGC..... 2524

41 ePProPheArgLeuArgThrAspProGluHisCysGlyArgArgGlyTyrG 58

2525 .CCTCCAAACCTGACTGCGGATGGCGCATGTGCGCTCTGGGAGACT 2573

58 IuLeuAspCysGlnAsnAsnGlnThrValPheAsnTyr..... 70

2574 CTTCCTGTGTGTGAGAACGATCTGCGCTGCGTAATTATGCTACATCAA 2623

70 70

2624 GGCCACTGCTACATCTCCAGACTCTGGGCTGACCCCATGTGCACCTG 2673

71LysSerArgIlePheAspValGlnIleLeuAsnT 82

2674 CCCCCCAGCCTTCACTGACAGCCGCTGCTC.....CTGGCTG 2711

82 YrArgSerTyrSer.....IleArgLeuLeu 90

2712 GGAACACCTTCACTCCACTGCTCAACCTGAACTCCCTTAAGAGTCACT 2761

91 AspProGlyLeuAsnAspGlnArgGluAsnGlyThrValPheProAsnH 107

2762 CAGCTTCTGCTCAGTAGAGAGAAATGCTCCATGAGGAGGCTCAACGC 2811

107 sArgAlaSerTyrAspAlaMetThrSerGlnIlePheGluThrValArgV 124

2812 CTCGGTGCATACAGACACTGGGAGACCTGGACATGCGGCGCTTCTCCGC 2860

124 alAsnAsnAspIleAsnTyrValAsnGlyLeuAlaProIleAsnSerS 140

2861 ..AACGCCAAGTGGACGAAATGCTGTGCAACCGCCCTGGGAAAGC 2908

141 GlnTyrIlePro.ThrSerPheCysSerLysAsnSerThrGlyPheSerT 157

2909 CCC..ATCCACACTGATGTGATCTGCGAGTCTCCAGT.....ACCG 2949

157 YrLeuValIleArgGluIleLeu..... 164

2950 CCTTCGGGCGCCGCTATTGACTTCTCGAACAACAGCTGCGCCGCGG 2999

165GlnAlaSerAspLeuAlaGlyCysArgValGluTh 177

3000 TGTGTGAGGGGTCTTATACACAGTTCACAGGAGAGAGAGAGCCAGG 3049

177 ValAlaTTPSerSerAlaProGlyIleSerSerLysSerSerThrL 194

3050 AACGAGCTGTCTTCCAGCCA.....TCTCCGAGAGAGAGCTCCGCGA 3093

194 eu.SerSerThrHisGlnGlyLeuAlaTyrGly...PheGluLeuSerT 209

3094 TGTGACAGCCCTGAACTGACGACGCTGAAGGCTTACTTCAATGCGATG 3143

```
209 pLysArgAsnLeu.....Leu. 214
|.....|
3144 GCTACAAAGGGCTACGACCTGGTCTACAGCCCAAGAGGGCTTACCTGC 3193
215 CysArgAsnCysAspArg.....SerArgIlyGlyIlyCysThrI1 228
|.....|
3194 GTGTCCCGGTGACAGTAGGGCTACGTGTACATGTAGAGGCGCATGCCAACA 3243
228 egluGluAsnSerAspArgAlaThrCys.....ArgT 239
|.....|
3244 CTTGCCCAAGTGGGCCCGGTGCAAGCTGTGTCTTCCTCCATCTACACGG 3293
239 yTrpCysIlyGluAspIleHisValSerIlyLeuThrPheArgCysIys 255
|.....|
3294 CTTGGGGCGAGCACTGTAGCACTGAC.....GCCCTGGG 3366
256 ValGluTyrTyrSerValTyrValLeuPhePheGlyIlyIleG1 272
|.....|
3323 ATGAACCTGACGCGCTTCTTCGCACTCTTGGG.....GCCCTGGG 3366
272 yGlyValLeuAla**ArgPheLeuGlyIlyLeuProIleLeuIleAla 289
|.....|
3367 CGGCGCTCTTG.....CTGCTGGGGGTCCGGAGCTTCGTGCTC 3404
289 lavalValTyrGlnCys 294
|.....|
3405 TGGGCTTCTGGGGTTCG 3421

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.AAS72520
seq_documentation_block:
ID AAS72520 standard; cDNA; 3875 BP.
XX
XX AAS72520;
AC
XX
XX 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #8324.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX MPI: 2001-639362/73.
XX
XX P-PSDB; ABG08333.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1: SEQ ID No 8324; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polypeptide chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
```

```
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
XX SQ Sequence 3875 BP; 824 A; 1213 C; 1060 G; 778 T; 0 other;
```

```
alignment_scores:
Quality: 95.00 Length: 341
Ratio: 0.605 Gaps: 20
Percent Similarity: 46.041 Percent Identity: 20.821
```

alignment_block:

us-09-805-919-1 x AAS72520 ..

Align seg 1/1 to: AAS72520 from: 1 to: 3875

```
29 CysValPro.....SerSerCysIlyAspIleGlnIleLysPph 41
|.....|
2496 TGGGCTTCTGGGAGAGGCGTGCAGGCGCTGC..... 2525
41 ePhePheArgLeuArgThrAspProGluHisCysGlyArgArgIlyTyrG 58
|.....|
2526 .CCTCCAAACCTGACCTGCGGATGGCGGACCTGCGCTGTGGAGACT 2574
|.....|
58 luleuAspCysGlnAsnAsnGlnThrValPheAsnTyr..... 70
|.....|
2575 CTTTCTGTGTGAGAACAGCTGCTGCTGTGATTAATTGCTACAAATCA 2624
70 ..... 70
2625 GGCCACTGCTACATCTCCAGACTGTGGCTGTACGCCATGTGCACCTG 2674
71 .....LysSerArgIlePheAspValGlnIleAsnT 82
|.....|
2675 CCCCCAGCCTTCACTGACAGCGCGCTGCTC.....CTGGCTG 2712
82 yTrgSerTyrSer.....IleArgLeuLeu 90
|.....|
2713 GGAACACTTCACTGCTCAACTGTCAACCTGAACTTCCTTAAGAGTCTC 2762
91 AspProGlyLeuAsnAspGlnArgGluAsnCysThrValPheProAsnH1 107
|.....|
2763 CAGCTCTTGTCTCAGTAGAAGAAATGCTCCATGAGAGAGTCAAGCGC 2812
107 sArgAlaSerTyrAspAlaMetThrSerGlnIlePheGlnTyrValArgV 124
|.....|
2813 CTGCGTGGCATACAGACTGGGAGCCCTGACATGGGGCCCTTTCGCGC. 2861
124 AlAsnAsnAspIleAsnTyrValAsnCysLeuAlaProIleAsnSerSer 140
|.....|
2862 ..AAGAGCCAGTGGAGAGATCATGTTGTGACAGCACCGCGCTCGGAAAGC 2999
141 GlnTyrIlePro.ThrSerPheCysSerLysAsnSerThrIlyPheSerT 157
|.....|
2910 CCC...ATCCAACTGATGTGCTATCTCGGAGTCCAGT.....ACCG 2950
157 yIleuValIleArgGluIleLeu..... 164
|.....|
2951 CCTCGGGGCCCGGTCAATTGACTTCTGTAACACACAGCTGCTGGCCGCG 3000
```



```

seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-616-844-7
seq_documentation block:
: Sequence 7, Application US/08616844
: Patent No. 5849578
: GENERAL INFORMATION:
: APPLICANT: FALB, DEAN A.
: TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE
: TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
: NUMBER OF SEQUENCES: 54
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/616,844
: FILING DATE: 15-MAR-1996
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/599,654
: FILING DATE: 09-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/485,573
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/386,844
: FILING DATE: 10-FEB-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: CORUZZI, LAURA A.
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7853-053
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6407 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
US-08-616-844-7

Alignment_scores:
Quality: 94.50 Length: 310
Ratio: 0.690 Gaps: 18
Percent Similarity: 44.194 Percent Identity: 23.226

alignment_block:
US-09-805-919-1 x US-08-616-844-7 ..

Align seg 1/1 to: US-08-616-844-7 from: 1 to: 6407
6 LeuLeuPheAlaCysValLeuLeuValPThrLeuIleSerSerLy 22
|||||::: ::: |||:::|||||
4099 CTGCGTCACTCTGAGCATCGCATGTATGTGTACCTTTCGAGAAAGAAATA 4148

```



```
4237 TGGGGCCGAGAACCTATGAAATGCAT...GAGAAATGGAAGTACC.... 4278
68 eAsnTYrLYsSerArgILEpheAspValGlnGluLeuAsnTYrARgSert 85
4279 .....AAAACCTCCTCCAGATGAGATGATGTACTGCGCCTA 4318
85 yISerILEarGLeuLeuAspProGLyLeuAsnAspGlnARgGlnAsnCYs 101
4319 CAAGTGTAAAG.....AATCCAGAACTTGAACGAAC 4350
102 ThValPheProAsnHisARgAlaSerTYrAspAlaMetThrSerGlnI 118
4351 GGACTCTACCCGGCTACACTGAGCTCCAGATCCAGCATCTCTGCAT 4400
118 ePheGluTPValARgValAsnAsnAspILEAsnTYrValAsnCYsLeuA 135
4401 TTTC..... 4404
135 labroILEAsnSerSerGlnTYrILEProThrSerPheCYsSerLYsAsn 151
4405 .....CCCGACAGATATAACCG...TCCTTATCATGATGTA 4440
152 SerThrGLyPheSerTYrLeuValILEarGLuILEuGlnAlaSerAs 168
4441 AGCAGAAAGAGAGACTACTTTAAAGTCAGAG..... 4473
168 PleuAlaGLyGLySARgValGluThrValAlaTrpSerSerAlaProG 185
4474 .....AGAGAGGAGTCAATGCTCTGACACCACTGACCTG 4507
185 lyLESerSerAsnLYsSer..... 191
4508 GGACCTCTGCTCAGAGACCGACCAAGAGGCTGGCCAGGATTTCG 4557
192 ...SerThrLeuSer...SerThrHisGlnGLyLeuAlaTYrGLyPheG 206
4558 GGAGCCACGCTGACTGCAAGCAAGCAAGAGGACAGGCTGGCGGCGTG 4607
206 uLeuSerTPeLYsArg.....AsnL 213
4608 ACCACAGTGGAGGAGACAGAGTGTGATGTGACACACAGGCTGCTCATG 4657
213 euLeuCYsARgAsnCYsAspARgSertARgLY..... 223
4658 CACCTTTGTGTACTGTGACTGATGATGTGGCCAGTATCAAGAGATC 4707
224 .....G 224
4708 TCTGTGAGTACTGCACCATGCATGCACACGAGGCGACTATGACCG 4757
224 lyGluCYsThrILEglnGluAsnSerAspARgAlaThrCYsArg.TYTr 240
4758 GGCAGACCACT.....AGACTTCAGTCCAGGAGACCTG 4789
240 pCYsLYsGLuAspILEHisValSerLYsLeuThrPheARgCYsLYsValG 257
4790 GTTTCCTCCTCGTTTGACATTGATAATTGGGTGGAGG.....TTTC 4833
257 luyTYrISerValTYrValLeuPhe 265
4834 CTYTTGATCTGTTTGAGACTGTTC 4859
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-944-868A-7
seq_documentation_block:
; Sequence 7, Application US/08944868A
; Patent No. 6018025
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,868A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/599,654
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: CORUZZI, LAURA A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-041
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 6407 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-944-868A-7

alignment_scores:
Quality: 94.50 Length: 310
Ratio: 0.690 Gaps: 18
Percent Similarity: 44.194 Percent Identity: 23.226

alignment_block:
US-09-805-919-1 x US-08-944-868A-7 ..
Align seg 1/1 to: US-08-944-868A-7 from: 1 to: 6407
6 LeuLeuPheAlaCYsValLeuLeuValThrLeuILESerSerLYs 22
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4099 CTGCTCATCTCCAGCACTCCGACATGATGTTGCTGCGAGAAATATA 4148
22 sAlaGlnAspILESerGlnCYsValProSerSerCYsGLyAspILEglnI 39
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4149 A...AATGACATACCAACTCATCTTCAAAAGT...GGAGATTCCAAA 4192
39 le.....lySPheProPheARgLeuARgThrAspProLuhIS 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4193 TGCCCCCATATGCTGAATACCCAAAATCTCGCTCA.....CAGAA 4236
52 CyasGLyARgARgLYrGLuLeuAspCYsGlnAsnAsnGlnThValPh 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4237 TGGGGCCGAGAACCTATGAAATGCAT...GAGAAATGGAAGTACC.... 4278
68 eAsnTYrLYsSerArgILEpheAspValGlnGluLeuAsnTYrARgSert 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4279 .....AAAACCTCCTCCAGATGAGATGATGTACTGCGCCTA 4318
85 yISerILEarGLeuLeuAspProGLyLeuAsnAspGlnARgGlnAsnCYs 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4319 CAAGTGTAAAG.....AATCCAGAACTTGAACGAAC 4350
```

```

102 ThrValPheProAsnHisArgAlaSerTyrAspAlaMetThrSerGlnI 118
      :::::||||| ::::: ||| |||
4351 GGACTGTACCGCGCTACACTGGACAGCCAGATACCGCATTTCTGCAT 4400
      |||
118 ePheGluTrpValArgValAlaAsnAspIleAsnTyrValAsnGlyLeuA 135
      |||
4401 TTTC..... 4404
135 LProIleAsnSerSerGlnTyrIleProThrSerPheCysSerTyrAsn 151
      :::::||||| ||| ||| |||
4405 .....CCCGACAGATTAACCC...TCTTTCACTCAGTATGAA 4440
152 SerThrGlyPheSerTyrLeuValIleArgGluIleLeuGlnAlaSerAs 168
      |||
4441 AGCAGAAGAAGAGACTACTTTTAAGTCCAGAG..... 4473
168 PheAlaIleGlyGlyCysArgValGluThrValAlaTrpSerSerAlaProG 185
      |||
4474 .....AGAGAGGAGCTCATTTGCTCTGAGCCAGTACCTG 4507
185 LylSerSerAsnLysSer..... 191
      |||
4508 GGACCTCTGCTCAGAGACCGCCAGAGAGGCTGCCGCCAGGATTTGTGCG 4557
192 ...SerThrLeuSer...SerThrHisGlnGlyLeuAlaTyrGlyPheG 206
      |||
4558 GGAGCCAGCGCTGAGTGCGAGAGAGGAGGAGGATGCGGGGCGCTG 4607
206 uLeuSerTrpLysArg.....AsnL 213
      |||
4608 ACCACAGTGGAGAGAGAGTGGAGTGGACACAGGCTGTCTATTGAG 4657
213 euLeuGlyArgAsnGlyAspArgSerArgly..... 223
      |||
4658 CAACCTTGTTGTACTGTGACGTGATGTGGCCAGATCAAGAGAGCTC 4707
224 .....G 224
4708 TCTCTGAGTACGACACATGGCAGCGACAGGCGCAGCTATTGCCAG 4757
224 LylGlyCysThrIleGluGluAsnSerAspArgAlaThrCysArgTyrTr 240
      |||
4758 GGCAGACCACT.....AGACTTCAGTGCAGGAGCGCTG 4789
240 PcyLysGluAspIleHisValSerLysLeuThrPheArgCysLysValG 257
      |||
4790 GTTTCCTTCGTTTGCACATTGAATATGGGTGGAG.....TTTC 4833
257 LuTrpTyrSerValTyrValLeuPhe 265
      |||
4834 CTTTGTGATCTGTTTGGAGCTGTTC 4859
seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-944-423A-7
seq_documentation_block:
; Sequence 7, Application US/08944423A
; Patent No. 6020463
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,423A
FILING DATE: 06-OCT-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,654
FILING DATE: 09-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/485,573
FILING DATE: JUN-07-1995
APPLICATION DATA:
APPLICATION NUMBER: US 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: CORUZZI, LAURA A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 6407 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-944-423A-7

alignment_scores:
Quality: 94.50 Length: 310
Ratio: 0.690 Gaps: 18
Percent Similarity: 44.194 Percent Identity: 23.226

alignment_block:
US-09-805-919-1 x US-08-944-423A-7 ..
Align seg 1/1 to: US-08-944-423A-7 from: 1 to: 6407
6 LeuLeuPheAlaCysValLeuLeuValThrIleuLysSerSerly 22
|||||: ::|||:|||||
4099 CTGCTATCTTACGACATGCCACTGATGTGTACCTTCCAGAAAGATTA 4148
|||:|||||: ::|||:|||||
4149 A..AATGACATAGCAAACTCATCTTCAAAGT...GGAGATTCCAAA 4192
39 le.....LysPheProPheArgLeuArgThrAspProGluHis 51
:::|||||: ::|||:|||||
4193 TGTCCCATATGCTGATACCCCAAAATCCGCTCA.....CAGAA 4236
52 CysGlyArgArgGlyTyrGluLeuAspCysGlnAsnAsnGlnThrValPhe 68
|||||: ::|||:|||||
4237 TGGGGCCGAGAAAGCTATTGAATGCAT...GAGATGGAGACTAC.... 4278
68 eaenTyrLysSerArgIlePheAspValGlnGluLeuAsnTyrArgSer 85
|||||: ::|||:|||||
4279 .....AAAACCTCTCCAGATGACGATGTGACTACGCGCTA 4318
85 yfserIleArgLeuLeuAspProGlyLeuAsnAspGlnArgLysCys 101
|||||: ::|||:|||||
4319 CAAGGTAGG.....AATCCAGACTTGAAGCAAC 4350
102 ThrValPheProAsnHisArgAlaSerTyrAspAlaMetThrSerGlnI 118
|||||: ::|||:|||||
4351 GGACTGTACCGCGCTACACTGGACAGCCAGATACCGCATTTCTGCAT 4400
118 ePheGluTrpValArgValAlaAsnAspIleAsnTyrValAsnGlyLeuA 135
|||||
4401 TTTC..... 4404

```

```

135 laProIleasnserSerIntYrIleProThrsPheCysSerLysAsn 151
4405 .....CCGGACAGTAATACCG...TCTTCATCAGTGAAGAA 4440
152 SerThngIyPheSerTyrlLeuValIleArgIuIleuGlnAlaSerAs 168
4441 AGCAGAGAGAGAGACTACTTTAATCCAGAG..... 4473
168 pheuAlaGlyGlyCysArgValGluThrValAlaIleTrpSerSerAlaProG 185
4474 .....AGAGAGGAGCTATTGCTCTGAGCCAGTCACTCG 4507
185 lyIleSerSerAsnLysSer..... 191
4508 GAGACCTCTGCTCAGAGAGACCGACAGAGGCTGCGCCAGATTGTGCG 4537
192 ...SerThleuser...SerThrlsGlnIyLeuAlaTyrlGlyPheG1 206
4558 GGAGCCACGCTGAGTGGCAGCAGAGAGAGGAGCAGCATCGGGGCGCTG 4607
206 uIeuserTrpLysArg.....AsnL 213
4608 ACCACAGTGGAGAGACAGAGTGAAGTGAACACACAGCTCTATTGAG 4637
213 eu.LeuCysArgAsnCysAspArgSerArgly..... 223
4658 CACCTTTGTTGTACTGTGACGTGAATGTGGCCAGTATCAAGAGATC 4707
224 .....G 224
4708 TCTCTAGTACTGCACCATGCGACTGGACACGAGCGCATTTAGCCAG 4737
224 lyGluCysThrIleGluGlnAsnSerAspArgAlaIleThrCysArg.TyrTr 240
4758 GAGGAGACACT.....AGACTTCAGTGCAGGAGACCTG 4789
240 pCysIySgIuAspIleHisValSerLysIleuThrPheArgCysLysValG 257
4790 GTTTTCCCTGCTTGGCACTTTAATAATGGTGGGAGG.....TTTC 4833
257 luTyTrSerValTyrlValLeuPhe 265
4834 CTTTGGATCTGTTTGGAGCTGTC 4859

seq_name: /cgm2_6/ptodata/2/ina/6A_COMB.seq:US-08-944-496-7

seq_documentation_block:
; Sequence 7, Application US/08944496
; Patent No. 6124433
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,496
; FILING DATE: 06-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654

```

```

; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6407 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-944-496-7

alignment_scores:
Quality: 94.50 Length: 310
Ratio: 0.690 Gaps: 18
Percent Similarity: 44.194 Percent Identity: 23.226

alignment_block:
US-09-805-919-1 x US-08-944-496-7 ..

Align seg 1/1 to: US-08-944-496-7 from: 1 to: 6407

6 LeuIleuPheAlaCysValIleuLeuValIleuIleSerSerly 22
|||||:.....: |||:|||||
4099 CIGCTCATCTGAGCATGCGACTGATTTGTCAGTTCAGAAAGATTA 4148
22 salagIuAspIleSerGlnCysValProSerSerCysGlyAspIleGln 39
4149 A...AATGACATAGCAAAATCATCTTCAAAAGT...GGAGATTTCGAA 4192
39 le.....LysPheProPheArgLeuArgThrAspProGlnHis 51
4193 TGCCCCCATATGCTGATACCCCAAAATCTCGCTCA.....CAGGAA 4236
52 CysGlyArgArgGlyTyrlGluLeuAspCysGlnAsnAsnGlnThrValPh 68
4237 TGGGGCCGAGAGACTATTGAATGCAT...GAGAAATGAGATACC..... 4278
68 eAsnTyrlYsSerArgIlePheAspValGlnGluIleAsnTyrlArgSert 85
4279 .....AAAACTCTCTCAGATGACGAGATGTACTACGCGCTA 4318
85 yIleIleArgLeuLeuAspProGlyLeuAsnAspGlnArgIuAsnCys 101
4319 CAAGGTGTAAG.....AATCCAGAACTTGAACGAAAC 4350
102 ThrValPheProAsnHisArgAlaSerTyrlAspAlaMetIleSerGlnI1 118
4351 GGACTCTACCGCGCTACACTGACTGCCAGATACGCACTTCTTGAT 4400
118 ePheGluTrpValArgValAsnAsnAspIleAsnTyrlAlaAsnCysLeuA 135
4401 TTTC..... 4404
135 laProIleasnserSerIntYrIleProThrsPheCysSerLysAsn 151
4405 .....CCGGACAGTAATACCG...TCTTCATCAGTGAAGAA 4440
152 SerThngIyPheSerTyrlLeuValIleArgIuIleuGlnAlaSerAs 168

```



```

4441 AGCAGAGAGAGAGACTACTTTTAAGCCAGAG ..... 4473
168 pleuhalglycylcysargvalglnthryalalatrpserealaprpg 185
4474 ..... AGAAGGAGACTATTTCTGTGAGCCAGTCACCTG 4507
185 lylseraserasnllyser ..... 191
4508 GGACTCTGTGTCAGAGAGACCGCACAGAGGCTGCGCCACAGATTTWGCG 4557
192 ..Serthrlreuser..Serthrlhslgnglyleualalatrplyphei 206
4558 GGACCCACGCTGATGGCAACAGCAGAGAGGACAGCATGGGGCGCTG
206 uleuSerTyrlysaarg ..... AsnL 213
4608 ACCAAGTGGAGGAGAGACAGGTGATGTGGAACACAGGCTGCTCATTCAG 4657
213 eu.Leucysalgaancysaspargseralrgly ..... 223
4658 CACCTTGTGTTCATTCTGTGAACGTGAATGTGGCCAGATCAAGAGAGTC 4707
224 .....G 224
4708 TCTCGATGATCACTGCACCATGTGACGTGGCAGGCGCATTTAGCCAG 4757
224 lyglucysThrIlegluglnasrAspargalatrhcysarg TyrTr 240
4758 GGCAGACCACT ..... AGACTCAGTGCAGGSACTGT 4789
240 pcyslysluaspIlehlslvalserlysluethrpheargCysValg 257
4790 GTTTTCCCTTGTTGACCTTTGATTAATGGGTGGAGG .....TTTC 4833
257 luTyrTyrserValTyrValleuphe 265
4834 CTTTGGATCTGTTTGAAGCTGTTTC 4859

seq_name: /cgn2_6/prodata/2/1na/6A_COMB.seq:US-08-434-000A-3
seq_documentation_block:
: Sequence 3, Application US/08434000A
: Patent No. 6046037
: GENERAL INFORMATION:
: APPLICANT: ANDREW C. HIATT, JULIAN
: APPLICANT: K.C.C. MA, THOMAS LEHNER
: TITLE OF INVENTION: IMMUNOLOGICALS CONTAINING PROTECTION
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: storage
: COMPUTER: IBM compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: Word Perfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/434,000A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA: including application
: PRIOR APPLICATION DATA: described below: 1
: APPLICATION NUMBER: 08/367,395
: FILING DATE: 12/30/94
: ATTORNEY/AGENT INFORMATION:

```

```

1 NAME: Guise, Jeffrey W.
2 REGISTRATION NUMBER: 34, 613
3 REFERENCE/DOCKET NUMBER: 212/127
4 TELECOMMUNICATION INFORMATION:
5 TELEPHONE: (619) 552-8400
6 TELEFAX: (619) 552-0159
7
8 TELEX: 67-3510
9
10 TELEX: SEQUENCE LISTING
11 INFORMATION FOR SEQ. ID NO.: 3:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 2919 base pairs
14 TYPE: nucleic acid
15 STRANDEDNESS: single
16 TOPOLOGY: linear
17
18 FEATURE:
19 NAME/KEY: Coding Sequence
20 LOCATION: 235....2472
21
22 US-08-434-000A-3

```

```
alignment_scores:          Quality:    92.50      Length:    304  
                          Ratio:       0.642     Gaps:        22  
Percent Similarity: 47.368 Percent Identity: 25.656
```

alignment_block:

Align seg 1/1 to: US-08-434-000A-3 from: 1 to: 2915

```

29 CysValProSer.....SerCysGlyAspIleGlnIleLysPheProPh 43
   |||||||
82 TGtGTCCCGAAAGTCAGATCAACGGAGGAGAGGAAGAGCGCTAAACATT 131
   |||||||
43 earGleuArgThrAspProGlnLysCysGlyArgArgGly.....56
   ::::|
132 GCACAGAGAAATCGCGCTGAGTGTGGCGCGCTGGGACCACACAGCAA 181
   ::::|
57 .....TyrIleuLysPspCysGlnAsnAsnGlnThrValPhe.....68
   |||||
182 TGCTGCTCTTCGTGCTCACTGCTCCG.....CTGGCGGTTTCCAGCC 225
   |||||
69 ...AsnTyrLysSerArgIlePheAspValGlnIleAsnTyr.....82
   ::::|
226 ATCTCCACAAAGATGCCCATATTGTGCTCCGAGGAGTGATTAAGTGGA 275
   ::::|
83 ...ArgSerTyrSerIleArgLeuLysAspProGlyLysAsnAspGln 98
   ::::|
276 AGGTAACTCATAGTGCATACGTGCTACTACCCACACCTCTTCACCC 322
   ::::|
98 rGlnLysnCysThrValPheProAsnHisArgAlaSerTyrAspAlaMet 114
   |||||
326 GG.....CACACCCGGAAGTAC.....342
   |||||
115 ThrSerGlnIlePheGlnIleTyrValArgLysAsnAspIleAsnTyrVa 133
   |||||
343 .....TGTGTCGGGAGGAGA.....GCTACAGC 365
   |||||
131 IaSnCysLeuAlaProIleAsnSerSerGlnTyrIleProThrSerPheC 148
   ::::|
366 TGCGTGCATAAACCCGTATGCTCTCGAGAGGCGTAGCTCCAGCAAAATAG 411
   ::::|
148 ySerLys.....AsnSerThrGlyPhe.....SerTyrLeuVal 155
   ::::|
416 CAGGCGAGGGCTAACTCCACCAACTTCCCGGAGAACGGACCACTTTGTGTG 465
   ::::|
160 IleArgGlnIleLeuGlnAlaSerAspLeuAlaGlyLysYsaArgValG 176
   ::::|
466 .....AACATGGCCAGGTGACGCCAGAGTAGTCCGGGC.....GCTA 503
   ::::|
176 IuThrValAlaIrrPserSer.....AlaProGlyIleSerSer 186
   ::::|

```

```

504 CAAGTGTGGCTGGGATCAATAGCCAGGCTGCTTTGATGATCAGCC 553
189 AsnlyserSer.....ThleuserSerThri 198
      :::::::::::|||||:|||||
554 TGGAGGTACAGCAGGCTCTGGGCTCTTAATGACACTAAGTCTACA.. 601
198 sGlnGlyLeuAlaTyrGlyPheGluLeuSerTyrPlyArg..... 211
      :::::::::::|||||:|||||
602 .....CAGTGGACCTGGGACAGAACGGTGACCAT 629
212 ..AsnLeuLeuCysArgAsnCysAspArgSerArgGlyGlyGlyThr 227
      :::::::::::|||||:|||||
630 CAATGCGCTTTCAGAGCTGAGATGCTCAAAAAGAGAGAGCTCTTACA 679
228 IleGluLysAsnSerAspArgAlaThrCysArgTyr...TyrCysArgS 243
      :::::::::::|||||:|||||
680 .....ACGATAGGCTGTACCTGTGCTGTGTCATCGAC 714
243 uAspIle.....HisValSerLysLeuThrp 252
      :::::::::::|||||:|||||
715 TCCAGTGTATGTGAATCCCACTATACAGAGAAATACGCTTGATAT 764
252 heArgCysValGlyTyrTyrSerValTyrValLeuPhePheGlyGly 268
      :::::::::::|||||:|||||
765 TCAGGCTACTGGCCAGTACTGTTCAGCGTGTGATCAACCAACTGAGGC 814
269 IleGlyIleGlyValLeuAla**ArgPheLeuLeuGly...IlePr 284
      :::::::::::|||||:|||||
815 TCACGATGCTGGGACGATATCTGCGCAGG.....CTGGGATGATTC 858
284 cIleLeuIle 287
      :::::::::::|||||:|||||
859 AATGATATA 868

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-312-157-3
seq_documentation_block:
/ Sequence 3: Application US/09312157
/ Patent No. 6303341
/ GENERAL INFORMATION:
/ APPLICANT: ANDREW C. HIATT, JULIAN
/ K.-C. MA, THOMAS LEHNER
/ TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
/ NUMBER OF SEQUENCES: 19
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: 633 West Fifth Street
/ Suite 4700
/ CITY: Los Angeles
/ STATE: California
/ COUNTRY: U.S.A.
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 MB
/ storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/312,157
/ FILING DATE: 14-May-1999
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/434,000
/ FILING DATE: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Guise, Jeffrey W.
/ REGISTRATION NUMBER: 34,613
/ REFERENCE/DOCKET NUMBER: 212/127
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 552-8400
/ TELEFAX: (619) 552-0159

```

```

?       TELEX: 67-351
?       SEQUENCE LISTING
?       INFORMATION FOR SEQ ID NO: 3:
?       SEQUENCE CHARACTERISTICS:
?         LENGTH: 2919 base pairs
?         TYPE: nucleic acid
?         STRANDEDNESS: single
?         TOPOLOGY: linear
?       FEATURE:
?         NAME/KEY: Coding Sequence
?         LOCATION: 235...2472
?         SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-312-157-3

alignment_scores:
  Quality: 92.50      Length: 304
  Ratio: 0.642      Gaps: 22
  Percent Similarity: 47.368      Percent Identity: 25.658

alignment_block:
US-09-805-919-1 x US-09-312-157-3

Align seq 1/1 to: US-09-312-157-3 from: 1 to: 2919

29 CysValProser.....SerCysGlyAspIleGlnIleLysPheProPh 43
      :::::::::::|||||:|||||
82 TGTTCCCTCAAGTCAGATCAACGAGAGAGAGAGAGTGGCTAAACATT 131
43 eArgLeuArgThrAspProGlnHisCysGlyArgArgGly..... 56
      :::::::::::|||||:|||||
132 GCACAGGAGAGAGTCGGCTGAGTGGCGGCTCGGACCCACAGCA 181
57 .....TyrGluLeuAspCysGlnAsnAsnGlnThrValPhe..... 68
      :::::::::::|||||:|||||
182 TGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 225
69 ..AsnTyrLysSerArgIlePheAspValGlnGluIleAsnTyr.... 82
      :::::::::::|||||:|||||
226 ATCTCCACGAGAGAGTCCCATATTTGTTCCCGAGAGAGTGAATAGTGA 275
83 ...ArgSerTyrSerIleArgLeuLeuAspProGlyLeuAsnAspGlnA 98
      :::::::::::|||||:|||||
276 AGGTACTAGTGTGCTACGCTGCTACTACCCACCACCTGTCACACC 325
98 rGluAsnCysThrValPheProAsnHisArgAlaSerTyrAspLamet 114
      :::::::::::|||||:|||||
326 GG.....CACACCCGGAAGTAC..... 342
115 ThrSerGlnIlePheGlnTyrPValArgValAsnAsnAspIleAsnTyr 131
      :::::::::::|||||:|||||
343 .....TGTTCCCGCAGAGGA.....GCTAGAGG 365
131 LasnCysLeuAlaProIleAsnSerSerGlnTyrIleProThrSerPhe 148
      :::::::::::|||||:|||||
366 TGGTGTGATTAACCTCATCTCCCGAGGCTACGCTCCAGAAATAG 415
148 ysSerLys...AsnSerThrGlyPhe.....SerTyrLeuVal 159
      :::::::::::|||||:|||||
416 CAGCAGAGGCTAACCTCACCACCTCCCGAGAGACGACATTGTGGTG 465
160 IleArgGluIleGlnAlaSerAsp..LeuAlaGlyGlyCysArgValG 176
      :::::::::::|||||:|||||
466 .....AACATGGCCAGCTGAGCCAGAGAGATCCGGGG.....GCTA 503
176 IuThrValAlaTyrPseSer.....AlaProGlyIleSerSer 188
      :::::::::::|||||:|||||
504 CAATGTGGCTGTGGCATCAATAGCCGAGGCTGCTTGTGATGTCAGCC 553
189 AsnlyserSer.....ThleuserSerThri 198
      :::::::::::|||||:|||||
554 TGGAGGTACAGCAGGCTCTGGGCTCTTAATGACACTAAGTCTACA.. 601

```

```

198 sglgIyLeuNlaTyrClYpHeLyuSeSerIPryLysArg..... 211
602 .....CAGTGGAACCTGGGCAACAAGCGTAGCAT 629
212 ..AsnLeuLeuCysArgAsnCysAspArgSerArgSglycyluCyStr 227
    :::||| ||| :::|||||::: |||||
630 CAACGTGCCCTTCACAGACTGAGAATGCTCAAAAGAGAGAACTCTTGTA 679
228 IleglUGluAseRsaPaRgaLaTrncCysAtgTy...TriPCysLySGl 243
680 .....ACCAATAGCCCTGTACCCTGTGCTGTCATGCAC 714
243 uAsPle.....HisValSerLySLeutRp 252
    ::||| ||| |
715 TCACAGTGGTATGTAATCCCAACTATACAGGACAGATACCGCTTGATAT 764
252 heargCysLyVaLIgLTyrSerValTyValLeuPhhehelly 268
    ||||| :::|||||::: |||||
765 TCAGGCTCTGCGCACGTACTGTTCAAGCGTGTGATCAACACACTCAAGC 814
269 IleglyIleglyValLeuaIa***ArgpheLeuLeugly...IlePr 284
    :::: ||| ::::||||| ||||| |||||
815 TCAGGAGTGTGGCAGCATCTCTGCGAG.....CTGGGATATATCC 858
284 oIleLeuTle 287
    |||||::||| |||
859 AATACTAATA 866

seq_name: /cogn2_6/prodata/2/lna/5A_COMB.seq;US-08-164-839-7

seq_documentation_block:
Sequence 7, Application US/08164839
Patent No. 5514573
GENERAL INFORMATION:
APPLICANT: YASUEDA, HISASHI
APPLICANT: NAKANISHI, KAZUO
APPLICANT: MOTOKI, MASAO
APPLICANT: NAGASE, KAZUO
APPLICANT: MATSUJI, HIROSHI
TITLE OF INVENTION: GENE ENCODING TRANSLUTAMINASE DERIVED
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: OHLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
City: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/164,839
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/004,729
FILING DATE: 14-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: ODION, NO. 5514573man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-599-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)412-3000
TELEFAX: (703)413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

```

```

?      LENGTH: 2088 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: double
?      TOPOLOGY: linear
?      MOLECULE TYPE: cDNA to mRNA
?      ORIGINAL SOURCE:
?      ORGANISM: Theragra chalcogramma
?      TISSUE TYPE: liver
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 1..2085
?      US-08-164-839-7

alignment_scores:
    Quality: 91.50      Length: 250
    Ratio: 0.855      Gaps: 10
    Percent Similarity: 42.800      Percent Identity: 21.600

alignment_block:
US-09-805-919-1 x US-08-164-839-7/review ..

Align seg 1/1 to reverse of: US-08-164-839-7 from: 1 to: 2088

86 SerIleArgLeuLeuAspProGlyLeuAsnAspGlnArgGluAsnCysrth 102
   |||::: ||| ||| |||:::
902 TGTGTGTATGGGCGTACCTCAAAAGTTGGTGGTATGACGACGGTGTGGGG... 858

102 rValPheProAsnHisArgAlaSerTyrTrpSplamethrSeriGlnIleP 119
   :::::|||||:::|||||:::
857 ...ATGCCAGGACATCGCAGCACTGTGCAGGCCACGGCTGCAACACC 812

119 heGluTrpValArgValaAsnAsp..... 127
   ::|||:: |||:: |||
811 ACATATGGCCACTACTCACTCCAGACTTGATTGGTGTCTCACTTCCACAG 762

128 .....IleasnTyrValaAsnCysLeuAlaProIleAsnse 139
   |||::: |||::: |||::: |||:::
761 ATGGGCACACTGCTCTCATTCATTGGTGGTCTTACTCCACCTGTGAAG 712

139 rSeriGlnTyrIleProThrSerPheCysSerLysAsnSerThrGlyPheS 156
   |||::: |||::: |||::: |||:::
711 TTCTCCCATTTTCCCAACG...ACTCCACCGTCATGTTGGCATCA 665

156 eTrrIleuValle.....Arg 161
   ::|||:: |||:: |||
664 CCATGGCAACGACGTGCTGATGTAGACAGAGGTCCATTCGCTGACGC 615

162 GlnIleLeuGlnAlaSerAspLeuIagly.....G1 172
   ::|||:: |||:: |||:: |||:: |||::
614 ATGTCTGTTGGGGGGCTGTACCTTGCTGTGCTTAAAGCGGTCAAGGATC 565

172 yGrsArgValGluThrValAlaTrpSerSerAlaProGlyIleSerSera 189
   |||::: |||::: |||::: |||::: |||:::
564 AAGACGATGTCATCATCTTGTCTTCAACAGTCCGAATTCATGCA 515

189 snLys..... 190
   ||
514 ACCCACTCATCCAAATCCGACATCCACAGTGAATGCGACATTCTCGTTC 465

191 .....SerSer.....ThrLeuSera 196
   |||::: |||::: |||::: |||:::
464 ATATATATATCTTGACGACACTTCTATATGAGGAGTACACAGATCAATC 415

196 rThrHisGlnGlyLeuAlaTyrGlyPheGluLeuSerTrpLysArgAsnL 213
   |||::: |||::: |||::: |||::: |||:::
414 TTGCACCGAGGATTTAAACAGCAAGTACATTTTG.....GTAAAGTCCG 371

213 euLeuLysArgAsnCys..... 218
   ::|||:: |||:: |||:: |||:: |||::
370 TCTCTCCCAAGATTTATTCCTGCAATACAGTTCAACACCAAGCGGTAC 321

219 .....AspArgSerArgGlyGlyGlyCysrThrIleGluAsnSeraS 233

```



```

APPLICANT: NAKANISHI, KAZUO
APPLICANT: MOTOKI, MASAO
APPLICANT: NAGASE, KAZUO
APPLICANT: MATSUI, HIROSHI
TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
FROM FISH
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSER: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/164,839
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/004,729
FILING DATE: 14-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: OBION, NO. 5514573man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-599-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)412-3000
TELEFAX: (703)413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2921 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Theragra chalcogramma
TISSUE TYPE: Liver
FEATURE:
NAME/KEY: CDS
LOCATION: 32..2122
US-08-164-839-9

alignment_scores:
Quality: 91.50 Length: 250
Ratio: 0.855 Gaps: 10
Percent Similarity: 42.800 Percent Identity: 21.600

alignment_block:
US-09-805-919-1 x US-08-164-839-9/rev ..
Align seg 1/1 to reverse of: US-08-164-839-9 from: 1 to: 2921
86 SerIleatgLeuLeuaspProglyLeuAsnaspGlnArgGluAsnCysTh 102
|||||
936 TGTGTGTCATGGCTGACTCAAAATGTTGATGACGCGTGTGGG..... 892
102 rValPheProAsnHisArgAlaSerTyrAspAlaMetThrSerGluIle 119
|||||
891 ...ATGCCAAGCATCCGACACTGTGCGACGCGCAGCGCTGCAAAACCC 846
119 heGltTIPValArgValAsnAsnasp..... 127
845 AGCATTGGCCATCTACTCATTCCAGACTTGTGATTGTCCTCAGCTTCCAGG 796

```

```

128 .....IleAsnTyrValAsnCysLeuAlaProIleAsnSe 139
795 ATGGGACACTGCTCATTCATTGGTGGTCTACTACCGCTGTAAAG 746
139 rSerGlnTyrIleProThrSerPheCysSerIysAsnSerThrGlyPhe 156
745 TTCTGCGCATTTCCCCACAG...ACTCCACCGTCATGTTGGATTCA 699
156 erTyrLeuValIle.....Arg 161
698 CCATGGCAGACTGCCCTGCTGATGTAGACAGGTCCTCATGACGC 649
162 GluIleLeuGlnAlaSerAspLeuAlaGly.....G1 172
648 ATGCGTTTGGGGGTGTGACCTGCTGCTGCTTAAAGCGTCAAGATCTC 599
172 yCysArgValGluThrValAlaTrpSerSerAlaProGlyIleSerSer 189
598 AAACAGATGTCATCATGATTGTCTCAAACTGTCGAAATTCATGCA 549
189 snIys..... 190
548 ACCCATCATCCAAATCCGACCTCCAGTGAATAATGACCATTCCTTC 499
191 .....SerSer.....ThrLeuSerSe 196
498 ATATGTAATTCCTTGACAGACACTTTCATCAGGAGGTACAGCATATC 449
196 rThHisGlnGlyLeuAlaTyrGlyPheGluLeuSerTrpIysArgAsnL 213
448 TTTCACAGAGGATTAACAGCAAGTACATTTTG.....GTAAGTCG 405
213 euleuCysArgAsnCys..... 218
404 TCCTCTCCAGAAAGTTATTCCTGACATCATTCACACCAAGCGGTAC 355
219 .....AspArgSerArgGlyGlyGlyCysThrIleGlnGluAsnSer 233
354 TCGCCAAGCGCAGCGGCGACGCTGTACACGATGACGATGATTCGTT 305
233 PArgAlaThrCysArgTyrTrpCysIysGluAspIleHisValSerIysL 250
304 CCTCGCGCTGTC.....TGTTGTAACCCACCTGTCACCAAGCCCTC 261
250 eutThrPheArgCysLysValGluTyrTyrSerValTyrValLeuPhe 266
260 CAGCCGTTCGCTGTGATCACAATATCATCATGTTGTTATCCCTGCTT 211

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-583-799-9

seq_documentation_block:
Sequence 9, Application US/08583799
Patent No. 5607849
GENERAL INFORMATION:
APPLICANT: YASUEDA, HISASHI
APPLICANT: NAKANISHI, KAZUO
APPLICANT: MOTOKI, MASAO
APPLICANT: NAGASE, KAZUO
APPLICANT: MATSUI, HIROSHI
TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
FROM FISH
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSER: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

598 AAGAGAGATGCCATCATCATATGTTCTTCAAACTGCCAAATTCATCATGCA 54
189 snUs..... 190
548 ACCCACTCATCATCCGCACTCCCACTGCAATGCAATGCACATTTCGTTG 49
191 .....SerSer.....ThleuSerSe 196
498 ATAATGATTCCTTGAGAGCACTTCATCATCAGGAGAGTACACAGCATATC 44
196 rThHisIcInclIyeuAiaIarIrgIpnHglIeuaSerTriplysArSaInL 213
448 TTTGACACAGGSAITTAACACCAAGTACATTTC.....GTAAAGTCG 405
213 eULeCySaArgASnCs..... 218
404 TCCCTCCGAGAGTTATTCCCTACATCATAGTTCACAGCCACGCGGTAC 35
219 .....AsparSerSerAGlyGLyGLucysThrIleGluGluAsnSerAs 233
354 TCGCCACAGGACGACGCGGACGAGCATGTACACATCAGCAATATTGTT 30
233 parGhIaThrCysaArgIyTrpCysLysGluAspIleHisIaSerIySL 250
304 CCTCCGCGCTGC.....TGCTGTAAACCACACTGTGCACACACGCTC 261
250 eULHPhaIRCySLysValGIuIyTrpSerValIyValIeupHaphe 266
260 CAGCGGTGCGCTGTGTGTATCATCATATATCGTTGTTTTCCTGCTT 211
seq_name: /cgn2.6/ptodata/2/lna/5A_COMB.seq:US-08-344-536-3
seq_documentation_block:
  Sequence 3, Application US/08344536
  Patent No. 5674735
  GENERAL INFORMATION:
    APPLICANT: Onions, David E.
    APPLICANT: Nicolson, Lesley
    TITLE OF INVENTION: EHV-4 GLYCOPROTEIN VACCINE
    NUMBER OF SEQUENCES: 5
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Bell, Seltzer, Park & Gibson
      STREET: Post Office Drawer 34009
      CITY: Charlotte
      STATE: No. 5674735th Carolina
      COUNTRY: USA
      ZIP: 28234
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/344,536
      FILING DATE:
        CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/961,672
        FILING DATE: 05-JAN-1993
        ATTORNEY/AGENT INFORMATION:
          NAME: Stibley, Kenneth D.
          REGISTRATION NUMBER: 31,665
          REFERENCE/DOCKET NUMBER: 1749-106
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: (919) 881-3140
          TELEFAX: (919) 881-3175
      INFORMATION FOR SEQ ID NO: 3:
        SEQUENCE CHARACTERISTICS:
          LENGTH: 1560 base pairs
          TYPE: nucleic acid
          STRANDEDNESS: single
          TOPOLOGY: linear
        MOLECULE TYPE: DNA (genomic)

```


FILING DATE: 14-JAN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: O'LOON, No. 55.4573man F.
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 10-559-C-
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)412-3000
 TELEFAX: (703)412-2220
 TELEX: 248653 OPTAT UR
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1921 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Theragra chalcogramma
 TISSUE TYPE: muscle
 US-08-164-839-28

```
alignment_scores:
  Quality: 89.50
  Ratio: 0.904
Percent Similarity: 42.857
Percent Identity: 20.779
```

```
alignment_block:
```

US-09-805-919-1 X US-08-164-839-28/rev

Align seg 1/1 to reverse of: US-08-164-839-28 from: 1 to: 1921

```

105 ProsnH1sArGalAserTYrAsPAlaMetHrSerGlnIlePheGluTr 121
||||| ||| :||| :||| :||| :|||
857 CCAGAGCATCCGATCAGCTGCTGACAGCCACGGCTGCMAACACCCAGATTG 808
121 pValArgValAsnAsp..... 127
807 GCCATTTCCTACTCCAGACTTGATTGTCTCCACTTTCAGGATGGCA 758
128 .....IlaSnTYrAlaSnGlyLeuAlaProIleAsnSerSerGln 141
757 CACGCTCATCATTTGGTTGGCTGATCCACCTCCAGCTGTGAAGTTCCTGC 708
142 TyrIleProHrSerPheGlySerLeuAsnSerThrGlyPheSerTyrl 158
707 CATGACCCACAGACAGATCCACCGCTATCATCGTGGCATTCACATGGC 658
158 vAlleArgLeu.....L 164
657 AACGACTCCCTGGATGATGACAGGAGTCCCATCGCTGAAGCATGGC 608
164 euGAlaSerAspLeuAlaGly.....GlycysArg 174
607 TTGGGGGGGTTCACCTCGCTGGTTAAAGCGGTCAAGATCTAAAGCA 558
175 ValGluTrValAlaIrrSerSerAlaPro..... 184
557 AGTCCATCATCATATCTTCAACTGTCGGAATTCATGCAACCACT 508
184 ..... 184
507 CATCCAAATCCGACATCCCAAGTAAATGCGACATTCCTCATATATG 458
185 ..GlyIleSerSerAsnIrrSerSer.....ThrIleSerSerThrHis 198
457 ATTCTTTAGACAGACATCATCAGGAGAGTACAGCGCATTTCTTTGCAC 408
199 GlnGlyLeuAlaTyrglyPheGluIleSerTrpLeuAsnLeuLeuGly 215
407 CAGGATTAACAGCAAGTACATTTTG.....GTAAAGCGGTGCTCTG 364
215 sArgAsnGly.....A 219

```

```

363 CAGAA GTTATTCCTCCATGACATGTTCCACAGCCAAAGCGGTA CTGCCCA 314
219 spArgSerArgGlyGlyGluCysThrIleGluGluGlnSerAspArgAla 235
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
313 CGGACAGCGGGCGGGAGCTGACACATCGACGGTAATGCTGGTTCGGC 264
236 ThrCysArgTyrTrpCysGlyAspIleHisValSerLysLeuThrP 252
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
263 CTCGCG.....TGGTGTACCCACACACTTGTACCGGACCTCCATCGT 220
252 eArgCysLysValGluTyrTyrSerValTyrValLeuPhePhe 266
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 TCGCTGTGTATCAACAAATATCATCGTGTATTATATCGTCTT 177

seq_name: /cgn2_6/prodata/2/1na_5A_COMB.seq:US-08-563-799-28

```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-583-799-28

```

seq.documentation.block:
Sequence 28, Application US/08583799
Patent No. 5607849

GENERAL INFORMATION:
APPLICANT: YASUEDA, HISASHI
APPLICANT: NAKANTSU, KAZUO
APPLICANT: MOTOKI, MASAO
APPLICANT: NAGASE, KAZUO
APPLICANT: MATSUI, HIROSHI
TITLE OF INVENTION: GENE ENCODING TRANSGUTAMINASE DERIVED
FROM FISH
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSER: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
City: Arlington
State: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/583,799
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/004,729
FILING DATE: 14-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NO. 5607849man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-599-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)412-3000
TELEFAX: (703)413-2220
TELEX: 248985 OPAT UR
INFORMATION FOR SEQ. ID NO.: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1921 base pairs
type: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Theragra chalcogramma
TISSUE TYPE: muscle
US-08-583-799-28

```

```
alignment_scores:      Length:    231
                        Quality:     89.50
                        Ratio:    0.904
                                Gaps:      8
```



```

76 AspValGInGluIleAsnTyrArgSerTyrSerIleArgLeuAsnAspPr 92
   |||||
152 TACGTCACAGAGATTAAGTATACCAACTTCTCATACAGGCTACGATCC 201
   |||||
92 oGlyLeuAsnAspGlnArgGluAsnCysThrValPheProAsnHisArgA 109
   |||||
202 CAGCCTGAAAAATCAGACCGAATAATGCTCTGTTTCCACACACAGCGG 251
   |||||
109 lAsSerTyrAspAlaMetThrSerGlnIlePheGluTrpValArgValAsn 125
   |||||
252 CAGGTACAGATGCGACGACGACGACGACGACGACGACGACGACGACG 301
   |||||
126 AsnAspIleAsnTyrValAsnCysLeuAlaProIleAsnSerGlnTyr 142
   |||||
302 AACGATATCAATTATATGTCACACTGTCGCGTCAATCAACTCTCATATA 351
   |||||
142 rIleProThrSerPheCysSer...LysAsnSerThrGlyPheSerTyrL 158
   |||||
352 CATCTCTACAGATTTTGTAGCACAGCTACAATCTCATGTTTGTAGCTATC 401
   |||||
158 euValIleArgGluIleLeuGlnAlaSerAspLeuAlaGlyGlyCysArg 174
   |||||
402 TTGTCGTAGGTAATAATTTGCAAGCTCAGATTAGCAGTGGTGGAGA 451
   |||||
175 ValGluThrValAlaIleTrpSerSerAlaProGlyIleSerSerAsnLys 191
   |||||
452 GTTGAAACTGTGGCAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 501
   |||||
191 rSerThrLeuSerSerThrHisGlnGlyLeuAlaTyr 203
   |||||
502 GTCTTCGTATCTCCAGCTTACCAAGCTCTGGTTAC 538

```

seq_name: gb_est2:BI434492

seq_documentation_block:

```

LOCUS      BI434492                639 bp    mRNA    linear    EST 21-MAY-2001
DEFINITION EST537253 P. infestans-challenged leaf Solanum tuberosum cDNA clone
PCBR09 5' sequence, mRNA sequence.
ACCESSION  BI434492
VERSION     BI434492.1 GI:15259182
KEYWORDS    EST.
SOURCE      potato.
ORGANISM    Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; easterids I; Solanales; Solanales; Solanales; Solanum.
            1 (bases 1 to 639)
REFERENCE   Restrepo, S., Griffiths, H.M., Smart, C.D., Cho, J., Chiemung, A.,
AUTHORS     Bougri, O., Buell, C.R., Ronning, C.M., Fry, M.E. and Baker, B.
            Generation of ESTs from potato leaves challenged with Phytophthora
            infestans, Compatible Interaction
            Unpublished (2000)
JOURNAL     Contact: Cathy Ronning
COMMENT     The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel: 1-800-711-6195, email cna@resgen.com
            Seq primer: M13F-R.
FEATURES
  Source
    1..639
    /organism="Solanum tuberosum"
    /cultivar="Kennebec"
    /db_xref="taxon:4113"
    /clone="PCBR09"
    /clone_lib="P. infestans-challenged leaf"
    /tissue_type="leaf"
    /dev_stage="6 week old"
    /lab_host="SOLR"
    /note="Vector: pBluescript SK(-); site_1: EcoRI; site_2:
    XhoI; Whole plants were challenged with 450,000
    sporangia/ml P. infestans US-1(US 940501) in Biotron
    (Madison, Wisconsin). Leaf tissue was collected at 1, 2,
    5, 12, and 24 hours post-challenge and frozen in liquid

```

nitrogen immediately upon removal). Kennebec plants showed no signs of HR. Katadin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."

```

BASE COUNT      200 a      144 c      101 g      194 t
ORIGIN
alignment_scores:
  Quality: 697.50      Length: 197
  Ratio: 4.079      Gaps: 4
  Percent Similarity: 86.802      Percent Identity: 69.543

```

alignment_block:
US-09-805-919-1 x BI434492 ..

Align seg 1/1 to: BI434492 from: 1 to: 639

```

8 PheAlaCysValLeuLeuLeuValThrLeuLeuSerSerLysAlaG 24
   |||||
47 TTGCTAGTGTCTTCTCTGTTTCATCTATACAGCTCTATAGCCCA 96
   |||||
24 nAsp...lIeSerGlnCysValProSerSerCysGlyAspIleGln...I 39
   |||||
97 AATATGTAAGTCAATGCTCAATGCTGCTTCTCTGTCGGAATTCGCCATA 146
   |||||
39 lElyPheProPheArgLeuArgThrAspProGluHisCysGlyArgArg 55
   |||||
147 TAAATTCCTCCCTCCGATGAGGACTGATCCTGAGCATGTGGCAACAT 196
   |||||
56 GlyTyrGluLeuAspCysGlnAsnAsnGlnThrValPheAsnTyrLys 72
   |||||
197 GAATGGAAGTCAATGCTGTCAGACAACTCAACCTTACACTTACAGCTC 246
   |||||
72 rArgIlePheAspValGlnGluIleAsnTyrArgSerTyrSerIleArgL 89
   |||||
247 CAGAAATTCCTACGTCGACGAAATTAATCAACAAGTACATGATGATGC 296
   |||||
89 euLeuAspProGlyLeuAsnAspGlnArgGluAsnCysThrValPhePro 105
   |||||
297 TACTGATCTGCTGCTTAAAGATCAGACGAAATTCGCTGTTTCCCT 346
   |||||
106 AsnHisArgAlaSer...TyrAspAlaMetThrSerGlnIlePheGluTr 121
   |||||
347 GATTACAGAGCAAAATTAATTAACGAGCTTAATGATGAAATTTCCAAAT 396
   |||||
121 pValArgValAsnAspIleAsnTyrValAsnCysLeuAlaProIleA 138
   |||||
397 GATTCATATTAACATAGCATCACTATGTCATGTCGAACTCTCTATCA 446
   |||||
138 snSerSerGlnTyrIleProThrSerPheCysSerLysAsnSerThr... 153
   |||||
447 ACTGTCACAAATACATTCACGAGCTTGTAGAACAAATACAACTTCG 496
   |||||
154 .....GlyPheSerTyrLeuValIleArgGluIleAsnAla 167
   |||||
497 CCCAAGCCAAATTTAGTATCTTCTTAAAGAAATATGCAAGCTTC 546
   |||||
167 rAspLeuAlaGlyGlyCysArgValGluThrValAlaIleTrpSerSerAla 184
   |||||
547 GCAATTTGGAACAGGTGCAAGGTGAAAGTGGCGCTCTACAGCTC 596
   |||||
184 rGlyIleSerSerAsnLysSerSerThrLeuSerSerThr 197
   |||||
597 CAGGTATTTTTCACAAACAGTGCCTTGTATGAGAGCACT 637

```

seq_name: gb_est1:AI897599

seq_documentation_block:

```

LOCUS      AI897599                533 bp    mRNA    linear    EST 18-MAY-2001
DEFINITION EST267042 tomato ovary, TAMU Lycopersicon esculentum cDNA clone

```

CLED28G17, mRNA sequence.
 A1897599
 A1897599.1 GI:5603501
 EST.
 SOURCE
 ORGANISM
 Lycopersicon esculentum
 tomato.
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 1 (bases 1 to 533)
 REFERENCE
 AUTHORS
 Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
 Liang,F., Upcon,J., Romning,C.M., Craven,M.B., Fujii,C.Y., Bowman
 ,C.L., Niernan,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
 ,S.D. and Giovannoni,J.
 Generation of ESTs from tomato carpel tissue
 Unpublished (1999)
 CONTACT: CUGI
 CLEMSON UNIVERSITY Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.
 FEATURES
 source
 1..533
 Location/Qualifiers
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="CLED28G17"
 /clone_lib="tomato ovary, TAMU"
 /tissue_type="carpel"
 /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
 /lab_host="XLI-Blue MRF"
 /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
 XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and
 directionally cloned cDNA in vector Lambda ZAP II with 5'
 and 3' ends located at the EcoRI and XhoI sites,
 respectively."
 BASE COUNT 135 a 103 c 120 g 175 t
 ORIGIN
 alignment_scores:
 Quality: 624.50 Length: 179
 Ratio: 4.055 Gaps: 2
 Percent Similarity: 86.034 Percent Identity: 66.480
 alignment_block:
 US-09-805-919-1 x A1897599 ..
 Align seg 1/1 to: A1897599 from: 1 to: 533
 97 GlnATGluAsnCysThrValPheProAsnHisArgAlaSerTyrAspAl 113
 ||| |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 2 CAGACGGAATTCGCTGCTGTTTCCACACAGGAGCAATTCAGATGC 51
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 113 aacethrserGlnIlePheGluTTPValAlaValAsnAsnAspIleasnt 130
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 52 CAtGATACCGCTTAATCTTCGATGCTCGTTCGATACCATATCAATT 101
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 130 yfValAsnCysLeuAlaProIleAsnSerSerGlnTyrIleProThrSer 146
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 102 ATGTCACTGTCGGCGTCAATCAATCACTCTCAGTATACATTCCTACAACT 151
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 147 pheCysSer..LysAsnSerThrGlyPheSerTyrLeuValIleArgG1 162
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 152 TTTTGTAGACACAGCTACAAATCTAGTTTGTCTTGTCTGAGGGA 201
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 162 uileLeuGlnAlaSerAspLeuAlaGlyGlyCysArgValGlnuhVala 179
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 202 AATATTGCAAGCTTCAGATTTCAGAGTCGTTCCAGAGATTGAAGCTGTGG 251
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 179 latrpserSerAlaProGlyIleSerSerAsnLysSerSerThrLeuSer 195

|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 252 CATGAGTGCATCTCCGGGAGATTCCACTACGATGCTCTGCTATCC 301
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 196 serThrHisGlnGlyLeuAlaTyrGlyPheGluLeuSerTyrPylAsnAs 212
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 302 AGCATTCACCAAGCTCTGCGATTATTCCTTCTGGAGCGCTGA 351
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 212 nleuLeuCySarGAsnCysAspArgSerArgGlyGlyGluCysThrIleG 229
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 352 TTTCTTAGCATAGAGTGT.....GGGAGAGAGAGAGAGATGTCTTCA 395
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 229 lngLAsnSerAspArgAlaThrCysArgTyrTrpCysLysGluAspIle 245
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 396 AAGACAATAGTATGTACCACTGTATCTACTCTGTAAAGAGACACT 445
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 246 HisValSerLysLeuThrPheArgCysLysValGluTyrTyrSerVal 262
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 446 CCGTTTCTGAAACGTTCTTGGATCGAATCGAAGTACATCTCTGTTT 495
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 262 rValLeuPhePheGlyIleGlyIleGlyIleGlyVal 274
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
 496 TGTATCGATCTATGTGTCATAGCAATGTGTGCACTT 532
 seq_name: gb_est1:A1490332
 seq_documentation_block:
 LOCUS A1490332 432 bp mRNA linear EST 18-MAY-2001
 DEFINITION EST248658 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
 CLED24P14, mRNA sequence.
 ACCESSION A1490332
 VERSION A1490332.1 GI:4385642
 KEYWORDS EST.
 SOURCE
 ORGANISM
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 1 (bases 1 to 432)
 REFERENCE
 AUTHORS
 Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
 Liang,F., Upcon,J., Romning,C.M., Craven,M.B., Fujii,C.Y., Bowman
 ,C.L., Niernan,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
 ,S.D. and Giovannoni,J.
 Generation of ESTs from tomato carpel tissue
 Unpublished (1999)
 CONTACT: CUGI
 CLEMSON UNIVERSITY Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 Location/Qualifiers
 1..432
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="CLED24P14"
 /clone_lib="tomato ovary, TAMU"
 /tissue_type="carpel"
 /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
 /lab_host="XLI-Blue MRF"
 /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
 XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and
 directionally cloned cDNA in vector Lambda ZAP II with 5'
 and 3' ends located at the EcoRI and XhoI sites,
 respectively."
 BASE COUNT 112 a 85 c 99 g 136 t
 ORIGIN
 alignment_scores:
 Quality: 516.50 Length: 145
 Ratio: 4.132 Gaps: 2
 Percent Similarity: 86.207 Percent Identity: 68.276


```

476 CTACAGAGA.....TACACTATTAGAGATTGAGAG...ATGGCAGT 516

166 aserapleualaglygysargvalglthrvalalatrpsersera 183
      ::::: |||||:::
517 CTCATAAGAGTGAATGATGACAGACAGAAATTATAGCTTCACCTCAT 566
      |||||:::
183 laprogllylesersersansyserserthleuserSerThHisglu 199
      |||||:::
567 GGCCCTATATATAC.....AACATTCCCTTCGTGATATTCATCA 607
      :::

200 GlyLeuAlaTyrglyPheGluDeuSerTrp 209
      ::::: |||||:::
608 GCAATTCCTACGAGATTGCTTCTTAT 637
      :::

seq_name: gb_est2:BE342777

seq_documentation_block:
LOCUS      BE342777                611 bp    mRNA    linear    EST 17-JUL-2000
DEFINITION EST395621 potato stolon, Cornell University Solanum tuberosum cDNA
clone GSTA21014, mRNA sequence.
ACCESSION  BE342777
VERSION     BE342777.1  GI:9252309
KEYWORDS    EST.
SOURCE      potato.
ORGANISM    Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1 (bases 1 to 611)
AUTHORS     Van der Hoeven, R., Bezzerides, J., Bachem, C., Horvath, B., Visser, R.,
            Holt, I.E., Liang, F., Hansen, T.S., Utterback, T., Bowman, C.L., Doan
            , B., Bougri, O., Buehl, C.R., Ronning, C.M., Tanksley, S.D. and Baker
            , B.
            Generation of ESTs from potato swelling stolons
            Unpublished (1999)
            Contact: Research Genetics, Libraries Division
            Tel.: 1-800-711-6195
            Email: cdna@resgen.com
            5 prime sequence.
FEATURES
    source          location/Qualifiers
                        1..611
                        /organism="Solanum tuberosum"
                        /cultivar="Bluntje"
                        /db_xref="taxon:4113"
                        /clone="GSTA21014"
                        /clone_id="potato stolon, Cornell University"
                        /tissue_type="axillary buds of stem explants, swelling
                        stolons"
                        /dev_stage="1 to 3 days"
                        /lab_host="SOLR"
                        /note="Vector: pBluescript SK(-) Site_1: EcoRI, Site_2:
                        XhoI; RNA was supplied by Christian Bachem & Beatrix
                        Horvath(Laboratory of Plant Breeding, Dept. of Plant
                        Sciences, Wageningen University, The Netherlands). Total
                        RNA was isolated from developing axillary buds of potato
                        nodal stem cuttings, cultured on medium for the
                        introduction of tuber formation as described in Bachem et
                        al. (Plant Journal 1996). Tissue samples were taken of
                        stages corresponding to growing stolons and the early
                        stages of tuber formation."
BASE COUNT  177 a      117 c      113 g      204 t
ORIGIN

alignment_scores:
    Quality: 288.00      Length: 187
    Ratio: 2.215      Gaps: 6
    Percent Similarity: 69.519      Percent Identity: 34.759

alignment_block:
US-09-805-919-1 x BE342777 ..
Align seg 1/1 to: BE342777 from: 1 to: 611

```

```

29 CysValProSerSerCysGlyAspIleGlu...IleTysPheProPhear 44
      ||| |||||:::
18 TGGTGCTCTTGTGCTTGTGGCATGTGGCTAAACATTAAGACCCCTTCA 67
      |||||:::
44 glnuargthrappproglunhiscysglylaryarglytyrtygluLeuasp 61
      ::|||:::
68 TTTAAACACCGAGATCCAAATATGTCATTTTCAGGATTAAGATTACCTT 117
      |||||:::
61 ysglnasnasnglnthrvalpheasntyrlysserargIlePheaspVal 77
      |||||:::
118 GGAAGAGTACAAACAGATTATAGCTATATGCTTCACCAAGAACTTCACGTT 167
      |||||:::
78 glnGluIleasnTyArgSerTySerIleargLeuLeuaspProGlyLe 94
      |||||:::
168 CAAAGCATAGACTATGATATACGAAATTCACCGTAGTCCGACTTT 217
      |||||:::
94 usnaspglnarGluasnCysThrValPheProasnhisargLa.... 109
      |||:::
218 ACAACACACAGATGATCTATGTCTCTCATACCTTCAAGCTTACCTTTC 267
      ||:::
110 ..SerTyrAspAlaMetThrSerGlnIlePheGlu.....TrpVal 122
      ::|||:::
268 TCAATATACAGTAGTTCTTCAAGCCGCTCATCAATATATACACTTGGTAT 317
      |||
123 ArgValAsnAsnAspIleasnTyArgValasnCysLeuAlaProIleasns 139
      |||
318 AGAAACCCGCCGCCATTTTCATGTTCAAAAGTCCATATGCTGTATATGA 367
      |||
139 rSerGlnTyrlleProThrSerPheCysSerTylsasnSerThrGlyPhe 156
      |||
368 TTCGACATTTGGAAATTAAGTGCGCTGCAATTAAGCAGA.....TTCA 411
      |||
156 erTyrlleValIleargGluIleleuGlnAlaSerAspLeuAlaGly 172
      ::|||:::
412 CTATTTAAAGATTGAGAA...ATGAAGTTTCGAAAGTAGAGARGGA 458
      |||||:::
173 CysArgValAlaThrValAlaTrpSerSerAlaProGlyIleSerSera 189
      |||||:::
459 TGCAGAGCAGAAATTATAGCTTCACCTCAGCAGCCCTATATTATAC.... 503
      |||||:::
189 nysSerSerThrleuserSerThHisGlnGlyLeuAlaTyrglyPheg 206
      ::|||:::
504 ...AACATTCCTTCGTGATTCATCAAGCAATTCCTACGAGATTGG 549
      |||||:::
206 luleuserTrp 209
      ||:::
550 AGTTTCTTAT 560

seq_name: gb_est2:B1922602

seq_documentation_block:
LOCUS      B1922602                725 bp    mRNA    linear    EST 17-OCT-2001
DEFINITION EST542506 tomato callus Lycopersicon esculentum cDNA clone
clone CUC78A19 5' end, mRNA sequence.
ACCESSION  B1922602
VERSION     B1922602.1  GI:16220301
KEYWORDS    EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE   1 (bases 1 to 725)
AUTHORS     Alcalá, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai
            , J., Utterback, T., Van Aken, S., Ronning, C.M., Fraser, C.M., Martin
            , G.B., Tanksley, S.D. and Giovannoni, J.
            Generation of ESTs from tomato callus tissue (2001)
            Unpublished (2001)
            Contact: CUGI
            Clemson University Genomics Institute
            Clemson University

```

100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 This clone is available through the Clemson University Genomics
 Institute
 Seq primer: T3.
 Location/Qualifiers
 1..725
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cLEC78A19"
 /clone_lib="tomato callus"
 /tissue_type="callus"
 /dev_stage="25-40 days old"
 /lab_host="XLI-Blue MRF"
 /note="Vector: pBluescript SK(-); Site:1: EcoRI, Site:2:
 XhoI; supplier: Boyce Thompson Institute; sequencing: The
 Institute for Genomic Research; cLEC - Cotyledons of
 seedlings 7-10 days post-germination were excised, cut at
 both ends and placed on MS medium with no selection. Mixed
 callus was harvested at 25 and 40 days and included
 undifferentiated masses. Tomato Callus EST Library"

BASE COUNT 205 a 148 c 128 g 244 t
 ORIGIN

alignment_scores:
 Quality: 274.50 Length: 293
 Ratio: 1.654 Gaps: 12
 Percent Similarity: 56.655 Percent Identity: 26.962

alignment_block:
 US-09-805-919-1 x B1922602 ..

Align seg 1/1 to: B1922602 from: 1 to: 725

```

6 LeuLeuPheAlaCyValLeuLeuValThrLeuIleSerSerSerly 22
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
25 CTTAGATATTTCACATTAATTCCTTAATTCATATTTCGCAACATCCAA 74
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
22 salaglnaspIleSerGlnCysValProSerSerCysGlyAspIleGln 38
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
75 TGTCTTGAGAGATCAAC..TGTCCGATTCTGCGTGGCGATTTCGCTA 121
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
39 ..lIeLysPheProPheArgLeuArgThrAspProGlnHisCysGly... 53
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
122 GCATAAATATACCTTTTCATTAACACTGATCCAAACATTTGTCGTCT 171
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
54 ArgArgGlyTyrGlnLeuAspCysGlnAsnAsnGlnThrValPheAsnTy 70
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
172 TTTTCAGAGATTGAATTAGCTTGTAAGGTAAACCAACCGTTATATGCTT 221
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
70 rIysSerArgIlePheAspValGlnGlnIleAsnTyrArgSerTyrSerI 87
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
222 ATCCCTCAGAGAGTTCATGATGTCAGAGCATGATGCTATATACACAA 271
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
87 leArgLeuLeuAspProGlyLeuAsnAspGlnArgGlnAsnCysThrVal 103
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
272 TTCAATTGGTAGATCCGACTTA...CAACACAGATCTATGCTCTCTC 318
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
104 PheProAsn.....HisArgAlaSerTyrAspAlaMetThrSerGlnI 118
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
319 CTACCTTCAAGCTTAACCTTCAACATATACGTAATTTCTCTGTCGATA 368
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
118 ePhe.....GluTPValArgValAsnAspIleAsnTyrValAsn 133
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
369 CTATTATGATCAGCGTATATGAGCCAGCCCATTTCTCATTTGTTCAAT 418
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
133 yLeuAlaProIleAsnSerSerGlnTyrIleProThrSerPheCysSer 149
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
419 GTCCATCTGTGTTGATGATTCGACATTTTGGAATATAGCGCGTC... 465
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
150 LysAsnSerThrGlyPheSerTyrLeuValIleArgGlnIleLeuGlnAl 166

```

```

466 ..AAATACCAATATACCTTACCTTAAGATTGGACA...ATGAAGT 509
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
166 aSerAspLeuAlaGlyCysArgValGluThrValAlaIlePheSerSera 183
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
510 CTCTCAAGTAGATGATGATGCAAGCGAGGTATATAGGTTAACCCTAT 559
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
183 lAProGlyIleSerSerAsnLysSerSerThrLeuSerSerThrHisGln 199
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
560 GCCCAATATTAAC.....AACATTCCATTTCTTAATTCATCAACA 600
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
200 GtIleuAlaTyrGlyPheGlnLeuSerTrpLysArgAsnLeuLysAsp 216
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
601 GCAATCTCTACGAGATTGACTTCCCTAT..... 630
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
216 gAsnCysAspArgSerArgGlyGlyGluCysThrIleGlnGlnAsnSera 233
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
631 .....TCTTCAGACTGGG..... 642
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
233 sPArgAlaThrCysArgTyrTrpCysLysGlnAspIleHisValSerLys 249
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
642 ..... 642
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
250 LeuThrPheArgCysLysValGluTyrTyrSerValTyrValLeuPheP 266
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
643 .....TTCTCCGCTGTTCCCGGTATG 668
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
266 eGlyGlyIleGlyIleGlyGlyValLeuAla**ArgPheLeuGlnI 283
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
669 GGGG.....GCCAAATTTGTGATAGGCC 691
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
283 lAProIleLeuIleAlaValAlaValTrp 292
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
692 TTCGCTGTGATATGCAATTCTGCGTAC 720
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

seq.name: gb-est2:B1922851

seq_documentation_block:

LOCUS B1922851 661 bp mRNA linear EST 18-OCT-2001
 DEFINITION EST542755 tomato callus Lycopersicon esculentum cDNA clone
 cLEC78A18 5' end, mRNA sequence.

ACCESSION B1922851
 VERSION B1922851.1 GI:16221050
 KEYWORDS EST.

SOURCE

ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

REFERENCE 1 (bases 1 to 661)

AUTHORS Alcala,J., Vredalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
 J., Uterback,T., Van Aken,S., Romling,C.M., Fraser,C.M., Martin
 G.B., Tanksley,S.D. and Giovannoni,J.
 Generation of ESTs from tomato callus tissue (2001)

JOURNAL Unpublished (2001)

CONTACT CUGI
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>
 This clone is available through the Clemson University Genomics
 Institute

Seq primer: T3.

Location/Qualifiers
 1..661

FEATURES
 source
 1..661
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cLEC78A18"
 /clone_lib="tomato callus"
 /tissue_type="callus"
 /dev_stage="25-40 days old"

SOURCE potato.
ORGANISM Solanum tuberosum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eusterids I; Solanales; Solanaceae; Solanum.
AUTHORS van der Hoeven R., Bezzerides J., Bachem C., Vissers R., Cho J., Chienango A., Bougri O., Buell C.R., Ronald C., Tanksley S. and Baker B.
TITLE Generation of ESTs from in vitro grown microtubers
JOURNAL Unpublished (2001)
COMMENT Contact: Cathy Ranning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@esgen.com
Seq primer: M13f-R.
FEATURES
Source
Location/Qualifiers
1..573
/organism="Solanum tuberosum"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone_lib="CST"
/clone_lib="CST"
/tissue_type="axillary buds of stem explants; growing sink-tubers"
/dev_stage="7, 8 and 10 days"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Tissue supplied by Christian Bachem and Richard Visser (Department of Plant Breeding, Wageningen University, The Netherlands). The CSTA libraries will attempt to capture the induction and initiation/initial growth of the tuber in an in vitro system as described in Bachem et al. (Plant Journal, 1996). Small microtubers develop from axillary buds attached to stem explants when placed on a high sucrose medium (10%). Visible morphological changes occur synchronously at day five in the axillary buds. The first library, CSTA (1-20) consists of axillary buds harvested on days 1-3. This targets those genes involved in induction of the microtubers. The following libraries, CSTA (21-40) and CSTA (41-60), capture genes involved in tuber initiation and outgrowth. This library is noted as P3 in Tanksley lab notebooks."

BASE COUNT 178 a 117 c 99 g 179 t
ORIGIN

alignment_scores:
Quality: 250.00 Length: 185
Ratio: 2.119 Gaps: 6
Percent Similarity: 63.784 Percent Identity: 34.595

alignment_block:
US-09-805-919-1 x B1179655 ..

Align seg 1/1 to: B1179655 from: 1 to: 573

```

6 LeuleuphealacysvalleuleuleuvalhrleuileserSerserly 22
||| : : : : : : : : : : : : : : : : : : : : : : : : :
29 CTTAATACTTACATTTACATTCCTTAACTCTTTCGCAACATCCAA 78
||| : : : : : : : : : : : : : : : : : : : : : : : : :
22 salaginaspiieserGincysvalproserSerserGysglvapllegin 38
||| : : : : : : : : : : : : : : : : : : : : : : : : :
79 TCGTTGGAAGATC...AGGTGCCGATTCCTTGGTCGCAATTCGCTA 125
||| : : : : : : : : : : : : : : : : : : : : : : : : :
39 ..llelysphepPhepbeaigleuarqhraspprogluhiscysgilyarg 54
||| : : : : : : : : : : : : : : : : : : : : : : : : :
126 ACAATAGACACCCCTTTTCAATTAACACGATCCAAACATTCGATTTT 175
||| : : : : : : : : : : : : : : : : : : : : : : : : :
55 ArggLYTYrGluLeuaspGysglvsnasnglnhrvalpheasnyrly 71
||| : : : : : : : : : : : : : : : : : : : : : : : : :
176 TANGACCTGATTTAGCTTGTAAGGTAAACCAACGATTATATGCTTATC 225

```

71 sSerArgillepheapvalgincguileasnyrArgserlyserliea 88
:
226 CTCACAAAGATTGCACGTTCAACACATGACATGATATATCAATCAAAATTC 275
:
88 rgleuleuaspProGlyleuasnaspGlnaagguasnGysrthrvalphe 104
:
276 ACCTGATGATCCGACTTACAAACACAAAGATGATATGCTTCTCATATA 325
:
105 Proasn.....HisArgAlaSerlyrAspAlaMethrSergl 117
:
326 CCTCTAGATTACCTTACACAAATACAGTAGTCTTCAGGGCGGCTGTA 375
:
117 ntlepheelutrpvalarg...ValasnaspIleasnTyValasnc 133
:
376 CTTATATACATCAGATCGTACGTAGACAGCCGCAATTTTCATGCTTACT 425
:
133 ValaAlaProIleasnserSerglnTyrlleProThrSerpheCysSer 149
:
426 GTCCATTGCGTGTATATGATTCACATTTGCGAAATATGCTGCTGCAAA 475
:
150 LysAsnSerThrGlyPheSerTyLeuValIleargGluileuglnal 166
:
476 CTAACCAAGA....TACACTATTATTAGAGATTGGAGAG...ATGGCAGT 516
:
166 aseraspLeuAlaGlyGysArgValGluThrValAlaTrpserSera 183
:
517 CTTAAAGTGAATGATGATGATGACAGACGAAATTTATAGGCTTCACCTCAT 566
:
183 lapro 184
:
567 GGCCCT 571

seq_name: gb_est2:BG455609

seq_documentation_block:
LOCUS BG455609 655 bp mRNA linear EST 19-MAR-2001
DEFINITION NF061D05P1F1044 Phosphate starved leaf Medicago truncatula cDNA
ACCESSION NF061D05P1 5', mRNA sequence.
VERSION BG455609
KEYWORDS BG455609.1 GI:13378934
SOURCE EST.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eusterids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
REFERENCE 1 (bases 1 to 655)
AUTHORS Liu D., Scott A.D., Harris A.R., Gonzales R.A., Bell C.J., Flores H.R., Inman J.T., Weller J.W., May G.D. and Harrison M.J.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula phosphate-starved leaf library
JOURNAL Unpublished (2000)
COMMENT Contact: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7325
Fax: 580 221 7380
Email: mharrison@noble.org
Insert length: 655 Std Error: 0.00
Plate: 061 row: D column: 05
Seq primer: TCACACAGCAACGCTATGAC.
FEATURES
Source
Location/Qualifiers
1..655
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_lib="NF061D05P1"
/clone_lib="Phosphate starved leaf"
/tissue_type="leaf"
/dev_stage="trifoliolate"
/note="Vector: Lambda Zap; At the trifoliolate stage, M. truncatula plants were transplanted to phosphate-free sand

and grown for a further 30 days. During this 30 day period, the plants were fertilized twice weekly with 1/2 Hoaglands solution containing only 200M potassium phosphate. RNA was prepared from above ground tissues."

BASE COUNT 210 a 141 c 114 g 189 t 1 others

alignment_scores:

Quality: 234.50 Length: 200
Ratio: 2.132 Gaps: 8
Percent Similarity: 55.000 Percent Identity: 33.000

alignment_block:

US-09-805-919-1 x BG455609 ..

Align seg 1/1 to: BG455609 from: 1 to: 655

```

2 LeuThrArgGlyLeuLeuPheAlaCys..... 10
   |||||
20 ATGACACAGAGGAGAAATCCTTACCTGTATGAGAAATCAATACAGCTT 69
   |||||
11 .....ValIleuLeuSerSerSerSerSerSerSerSerSerSerSer 14
   |||||
70 TAACTCTCTCCAAATTGTGACNGGTGTTACTTGTGTACTATTCATC 119
   |||||
15 .....ValIleuLeuSerSerSerSerSerSerSerSerSerSerSer 28
   |||||
120 ATCATCATCAACATGATGATGATGATGATGATGATGATGATGATGAT 169
   |||||
29 CysValProSerSerSerSerSerSerSerSerSerSerSerSerSer 44
   |||||
170 TGTCCACCTCTTCATCGCGCAAAATACAAACATATCATCATCTTTCG 219
   |||||
44 GlnuArgThrAspProGlnuHisCysGlyArgArgGlyTyrGlnuAsp 61
   |||||
220 ACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 269
   |||||
61 YsgIAsnAsnGlnThrValPheAsnTyrIysSerArgIlePheAspVal 77
   |||||
270 GCGAAACACATCATCCTGTTAACTCTGTTTCAAGTAAATGATGATG 319
   |||||
78 GlnGluIleAsnTyrArgSerTyrSerIleArgIleuAspProGlyLe 94
   |||||
320 CAGGAAATCACTAGCTAAATTCACAAATTCGAGTGTGATCCAGTAAT 369
   |||||
94 UbsnAspGlnArgIuAsnCysThrValPheProAsnHisArgAlaSer 111
   |||||
370 T.....GAGAGGGTAATGTCTCTCATTCCTCATATTTCTTACTA 413
   |||||
111 yTAspAlaMetThrSerGlnIlePheGlnTTPValArgValAsnAsp 127
   |||||
414 AATCCAAATTCACAAAGT.....TCTTCAATTCATATCAACAAAGAG 454
   |||||
127 ..... 127
455 GATCCATATCAATATGATGATGATGATGATGATGATGATGATGATG 504
   |||||
128 .....IleAsnTyrValAsnCysLeuAlaProIleAsnSerSer...GlnT 142
   |||||
505 GCACATTAATATATCTGAAATGTGACAAACAAGTGAATGATGATGATG 554
   |||||
142 yTlleProThrSerPheCysSerIysAsnSerThrGlyPheSerTyrLeu 158
   |||||
555 ATGTGATATCAGCTCCCTGTATC...AACTCCGACTCCGAAAGCTATCTT 601

```

seq_name: gb_est1:AW560216

seq_documentation_block:

LOCUS AW560216 547 bp mRNA linear EST 07-SEP-2000
DEFINITION EST315264 DSR Medicago truncatula cDNA clone pDSIR-26012, mRNA
sequence.
ACCESSION . AW560216

VERSION

AW560216.1 GI:7205642

KEYWORDS

EST.

SOURCE

barrel medic.

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

REFERENCE

1 (bases 1 to 547)

AUTHORS

Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Peng

TITLE

H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,

JOURNAL

ESTs from roots of Medicago truncatula after inoculation with

COMMENT

unpublished (1999)

JOURNAL

Contact: Carroll P. Vance

COMMENT

Department of Agronomy and Plant Genetics

JOURNAL

University of Minnesota

COMMENT

411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA

JOURNAL

Tel: 612 625 5715

COMMENT

Fax: 651-649-5058

JOURNAL

Email: Vance004@maroon.tc.umn.edu

COMMENT

Minnesota EST name:M251571e ; TIGR sequence name:MTBAV90TK ; More

JOURNAL

information, including clone ordering, is available at.

COMMENT

'http://chrystle.tamu.edu/medicago'

JOURNAL

Seq primer: SKmod (CPA gAA CTA gta gAT CC).

COMMENT

Location/Qualifiers

FEATURES

source

location

qualifiers

base

count

origin

alignment_scores:

Quality: 233.50 Length: 179
Ratio: 2.245 Gaps: 7
Percent Similarity: 58.101 Percent Identity: 33.520

alignment_block:

US-09-805-919-1 x AW560216 ..

Align seg 1/1 to: AW560216 from: 1 to: 547

BASE COUNT 164 a 122 c 80 g 181 t

ORIGIN

```

28 GlnCysValProSerSerCysGlyAspIle...GlnIleYspheProph 43
   |||||
17 CATCTTTCTTCTCTCTCTGAGAAATTTGCAATTAATCATTCATTCAT 66
   |||||
43 eArgLeuArgThrAspProGlnuHisCysGlyArgArgGlyTyrGlnuAs 60
   |||||
67 CCGACTAAACACAGCAGCAGCAGCTGTGGCAACAAATTTGACAGATGGA 116
   |||||
60 spCysGlnAsnAsnGlnThrValPheAsnTyrIysSerArgIlePheAsp 76
   |||||
117 ATTGGCTAAACACAGTTCAGTATGAAAGCTGTTGATGCTGAATATTTT 166

```

(/note=Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; RNA was supplied by Christian Bachem & Beatrix Horvath(Laboratory of Plant Breeding, Dept. of Plant Sciences, Wageningen University, The Netherlands). Total

```

alignment_scores:      Quality:      233.00      Length:      158
                        Ratio:      2.284      Gaps:      3
Percent Similarity:    64.557      Percent Identity: 32.911

alignment_block:
US-09-805-919-1 x AW906806      ..

Align seg 1/1      to: AW906806      from: 1      to: 543

```

[illegible]

Asteridae; eusteriids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 1 (bases 1 to 710)
 Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
 J., Uteback,T., Van Aken,S., Romling,C.M., Fraser,C.M., Martin
 G.B., Tanksley,S.D. and Giovannoni,J.
 Generation of ESTs from tomato callus tissue (2001)
 Unpublished (2001)
 CONTACT: CUGI
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 This clone is available through the Clemson University Genomics
 Institute

Seq primer: T3

FEATURES

Source Location/Qualifiers

1..710
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone_lib="CLEG78N4"
 /clone_lib="tomato callus"
 /tissue_type="callus"
 /dev_stage="25-40 days old"
 /lab_host="X11-Blue KRF"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; supplier: Boyce Thompson Institute; sequencing: The
 Institute for Genomic Research; cLEC - Cotyledons of
 seedlings 7-10 days post-germination were excised, cut at
 both ends and placed on MS medium with no selection. Mixed
 callus was harvested at 25 and 40 days and included
 undifferentiated masses. Tomato Callus EST Library"

BASE COUNT 204 a 158 c 131 g 217 t

ORIGIN

alignment_scores:

Quality: 232.00 Length: 165
 Ratio: 2.071 Gaps: 6
 Percent Similarity: 67.879 Percent Identity: 35.152

alignment_block:

US-09-805-919-1 x B1922909 ..

Align seg 1/1 to: B1922909 from: 1 to: 710

```

31 ProSerSerCysGlyAspIleGln..IleYsPheProPheArgLeuAr 46
|||||.....| | | | | | | | | | | | | | | | | | | | | |
108 CCTTCTGCTGTGGTCATATCCGACATCAGCTACCTTTTCACTTAA 157
46 gThrAspProGluHisCysGly...ArgArgGlyTyrgLuleuAspCysG 62
|||||.....| | | | | | | | | | | | | | | | | | | | | |
158 CAGTGACCCAGAAATTTGTGAGAGATCCGAAATTTGAATTTAGTTGTG 207
62 InAsnAsnGlnThrValPheAsnTyrgLysSerArgIlePheAspValGln 78
|||||.....| | | | | | | | | | | | | | | | | | | | | |
208 AAGATACCAACGCTTATGTCATCTCTCCAGAGAGCTGTATGTCCA 257
79 GluIleAsnTyrgLysSerTyrgLysIleArgLuleuAspProGlyLeuAs 95
|||||.....| | | | | | | | | | | | | | | | | | | | | |
258 GCCATCACTATTAATAGTACGACATTCACCTGTAGATCCAGCTTTACA 307
95 nAspGlnArgGluAsnGlyThrValPheProAsn.....HisArgAlas 110
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
308 AAGCAACAAGATGATCTATGCTCTTCAAGTCCACACTCTCTCTCGACC 357
110 eTyrgAspAlaMetThrSerGlnIlePheGluTyrgValArgValAsnAsn 126
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358 AATCCGATACCAATCTCCGATCATCATATAGTGGCTTAGATCAGCAGAG 407
127 AspIleAsnTyrgValAsnGlyLeuAlaProIleAsnSerSerGln...Ty 142
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

408 CCCATTTTCATGTTCAACTGTCATCTGCTGTTAATAGTTCTTCGACATT 457
142 rIleProThrSerPheCysSerLysAsnSerThrGlyPheSerTyrgLeuV 159
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
458 TCTGGAATTAAGTGGCTGCAAAATTAAGCAGG.....TACACTTATTTA 501
159 aIleArgGluIleLeuGlnAlaSerAspLeuAlaGlyGlyCysArgVal 175
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
502 AGATTGAGAA...ACGAAAGTCTGTGAGTCAAGCCATGATGCAAGATG 548
176 GluThrValAlaTrpSerSerAlaProGlyIleSerSerAsnLys 190
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
549 GAATTATAGCCATGACCTCATTCCTGCTGATATTAACGACAGAGA 593

```



```

1145 AACACCAAGCTGCTGCTTCTTCTGCTGAGCATCTTGAAGTACGGT 1096
41 heProPheArgLeuArgThrAspProGluHisCysGlyArgArgGlyTyr 57
1095 ACCCTTCCTCGTTACAGGTGATCCACCTGCTGGTGAATGAAGCTAT 1046
58 GluLeuAspCysGlnAsnAsnGlnThrValaHeuSerTyrLysSerArg 74
1045 GAGCTAGTCTGTAGAGATGGACAGCATATATTCATCAACACAGGAAA 996
74 eHeuAspValGlnGluLeuAsnTyrArgSerTyrSerLileArgLeu 91
995 ATATTGTGACAGACATCTCTCTATACAGATCTAGATTTGGTGTG 946
91 sPProGlyLeuAsnAspGlnArgGluAsnCysThrVal...PheProAsn 106
945 ATGCTAATTTAGACAAAC...AGCACTTCCCTCTCCACGTGGGAA 902
107 HisArgAlaSerTyrAspAlaMetThrSerGlnLeph... 119
901 AATCTTCCCTACTTCATGACATGGGCACTAGCTTACACTTCTGCACT 852
120 .GluTyrValaArgValaAsnAsnAspLleAsnTyrValaAsnCysLeuAla 136
851 TCGGTGGGCT...ACATTTCTAATTTGCTCACGGG 820
136 roLleAsnSerSerGlnTyrLileProThrSerPheCysSerLysAsnSer 152
819 CATTAACAAAGCATGTACATGCTGCTGCT...TGGTGAAGGGGAAT 773
153 ThrGlyPheSerTyr...LeuValLleArgGluLileuGlnAlaSer 168
772 ACTCTTTGTTATGCTGTGACATGACAGCTCTTACTATGTTCAAA 723
168 PleuAlaGlyCysArgValaGluThrValaLleTyrSerSerLarPro 185
722 CATGAGCCTTCTGTGATCTGCTGGCTG...ATTCTG 685
185 LylleSerSerAsnLysSer...SerThrLeuSerSerThr 197
684 TGGATGATCGTACAAAGATGACACACTGACAGATGACAGATG 635
198 HisGlnGlyLeuAlaTyrGlyPheGluLeuSerTyrLysArgan... 212
634 GTGAATTCATGAGGAATGTTTCTGCTTCTGCTGAGCTTCTT 585
213 .....LeuLeuCysArgAsnCysA 219
584 GTTGTATCTATTATAAAATATATATCTCTTTGCAATCTGCGAG 535
219 sParGSerArgGlyGlyLucysThrLleGlu.GluAsnSerAspArgAl 235
534 AAACCTTTAAG...AGCAAAATGT...ATAGTAGAAGACATTCMAAACTGA 491
235 aThrCysArgTyrTyrCysLysGluAspLleHisValSerLysLeuThr 252
490 CTTTCGCCATTTATGGACCGAGCTACAAAT...T 459
252 heArg.CysLysValaGluTyrTyrSer...Va 261
458 ITAGGATGTGTATTAATTAATTAATCTAGTCCACGCAAGTTGGCTTAC 409
261 lTyrValleuPhePheGlyGlyLleGlyLleGlyValaLeuAla**A 278
408 TCTCGTTCTGTATTT...GCCCTAGAGATTGTTA 377
278 rGpHeuLeuGlyLleProLleLeuile 287
376 AGTGCATTTATAGTACTCCAGATATGCTT 348
seq_name: gb_p1:AF238476

```

```

seq_documentation_block:
LOCUS      AF238476                3628 bp    DNA       linear   PLN 21-JUN-2000
DEFINITION Oryza sativa receptor-like kinase (RLG1) gene, complete cds.
ACCESSION  AF238476
VERSION    AF238476.1 GI:18575485
KEYWORDS
SOURCE     Oryza sativa.
ORGANISM  Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE  1 (bases 1 to 3628)
AUTHORS   Yun,C.-H., Lee,G.-R. and Kim,H.-I.
TITLE      Molecular analysis of receptor-like kinase in rice
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 3628)
AUTHORS   Yun,C.-H., Lee,G.-R. and Kim,H.-I.
TITLE      Direct Submission
JOURNAL    Submitted (24-FEB-2000) CytoGenetics, National Institute of
FEATURES   Agricultural and Technology, Seodundong 249, Suwon 441-707, Korea
SOURCE     1..3628
            /organism="Oryza sativa"
            /variety="Nipponbare"
            /db_xref="taxon:4530"
            /chromosome="1"
            join(<1327..2066,2348..>3623)
            /gene="RLG1"
            /product="receptor-like kinase"
            <1327..>3623
            /gene="RLG1"
            join(1327..2066,2348..3623)
            /gene="RLG1"
            /note="protein kinase"
            /codon_start=1
            /product="receptor-like kinase"
            /protein_id="AAF78020.1"
            /db_xref="GI:8575486"
            /translation="MHPLSTWQAFVFFFAVLALLAGVDSGRCYFPGSGLHNTSH
            PERRRDLPORCPVSELDRCDSKATIRINGTGYVRSINVTTFVFWVDSLKDRNS
            SCPLRSDLPVSGIIOGSHGMDLADPGATWFSVNSQAVNNVSVYVDTL
            STSSSVYVFGSWIMPSVFPFICGLTETSCRYAMIPGSGDPSLPNNSFSDIVSM
            RNFANHPPIIHRMSRIGHIKDCLMGSRIIRIHICTICFCFVDCWYDYSEYLI
            QITITAPPOCKTHLKVIVYVAVCRVAVLPLTLFLFKYWKYIKID
            AVEKFLQMLMGPRTAYATDIIAMTSFRKLOGGGVKCVILPGDVHAIKIL
            SNYCNGBEIEFIEVSTIGSIHHVNVVRVGYCAEMRSALVEYMPGSDRFESFD
            KSLMDKLEIALGIARGINYLHOGCDQIILHPDIKPHNILLDSNFYKADGILAK
            YPDNSFEMVSAARGVGYIAPEMISRSFGIISKSDVYSGLMLLMAGGRRSKON
            MSSSSQSYRPSMVYNOIYOOKGRTANAFNNHELEKICVYGLICIQMKSHDRPTKE
            VTEMESDVGGIQLBSRPFCDDEPLPLIYVDSRFSSELTISEDE"
BASE COUNT      921 a      738 c      869 g      1080 t
ORIGIN
alignment_scores:
Quality: 193.00      Length: 243
Ratio: 1.451      Gaps: 10
Percent Similarity: 54.733      Percent Identity: 25.926
alignment_block:
US-09-805-919-1 x AF238476
Align seg 1/1 to: AF238476 from: 1 to: 3628
7 LeuPheAlaCysValLeuLeuLeuValThrLeuLleSerSerLysAl 23
1366 ATCTTGCTGTTGCTGCTCTTCTGACAGAGATTCGAGGGGGCGGCAAG 1415
23 acLInAspLleSerGlnCysValProSerSerCysGlyAspLleGln...I 39
1416 G.....TGCATCTCTTCTCTCTGTCGACATCTCCAGACA 1450
39 lelysPheProPheArgLeuArgThrAspProGluHisCysGlyArgArg 55

```



```

121 .....TrpV 122
2506 ACATGGCGGCTGGGACCTGACCCCTGACCTGAGCTGACATGCG 2457
122 alarValAsnaAsnAspIleAsnTyrValAsnCysLeuAlaProIle... 137
2456 TT.....ACCTTGTGCAATTGTTTCACAGCAGTAAGG 2425
138 AsnSerSerGlnTyrIleProThrSerPheCysSerLysAsnSerThrG1 154
2424 AATAAATAGTGTGACCTCTCTGTGAT...TCCCTGAGGACACAGCTCTTC 2378
154 yPheSerTyrLeuValIleArgGluLeuGlnAlaSer..... 167
2377 GTTGTGTATGTTGTTGGTCTCGATATAGCTCCATCTGTTTCCAA 2328
168 .....AspLeuAlaGlyGlyCysArg.....ValGluThr 177
2327 TCATCGGAACCTTGAGACTTCATGCGTACCTGACATGATTCCTTG 2278
178 ValAlaTyrSerSerAlaProGlyIleSerSerAsnLysSerSerThrLe 194
2277 GGTGGTGGGACTCGCCATTGCCACATATGCAACCTTTTCAGACATCGT 2228
194 userSerThrHisGlnGlyLeuAlaTyrGlyPheGluLeu.....SerT 209
2227 CAGATCGATGAGGAGATGATGATTGCTGTTGTTTCCATATACATAGT 2178
209 rPlySarGAsnLeuLeuCysArgAsnCys 218
2177 GGAGTAGGATTGGGACATCAAGGATTGT 2149
seq_name: gb_pl:AP003219

```

```

seq_documentation_block:
LOCUS      AP003219          126434 bp      DNA      linear      PLN 03-JUL-2001
DEFINITION      Oryza sativa genomic DNA, chromosome 1, BAC clone:OSJNB0032H19.
ACCESSION      AP003219
VERSION
KEYWORDS
SOURCE
ORGANISM      Oryza sativa (cultivar:Nipponbare) DNA, clone:OSJNB0032H19.
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
AUTHORS      1 (bases 1 to 126434)
              Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE      Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
              clone:OSJNB0032H19
JOURNAL
REFERENCE
AUTHORS      2 (bases 1 to 126434)
              Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE      Published Only in Database (2001) In press
JOURNAL
REFERENCE
AUTHORS      Direct Submission
              Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
              Agricultural Resources, Rice Genome Research Program; Kannondai
              2-1-2, Tsukuba, Ibaraki 305-8602, Japan
              (E-mail:tsasaki@ar.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
              Tel:81-298-38-7441, Fax:81-298-38-7468)
              On Jul 3, 2001 this sequence version replaced gi:13620984.
              GENS CAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
              (October 1998 version). The genomic sequence was searched against
              NCBI NonRedundant Protein database, nr
              (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
              RGP. Protein homologues of the coding regions were searched against
              NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
              the identified cDNA sequences using BLASTN 2.0 with the
              corresponding DBJ accession no. and RGP clone ID.
              A gene with identity or significant homology to a protein is
              classified based on the protein name to indicate the homology level
              such as same name, 'putative-' and '-like protein'. A gene without
              significant homology to any protein but with EST homology (covering

```

FEATURES

almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein. The orientation of the sequence is from M3rev to -21M13 of the BAC clone. This sequence of OSJNB0032H19 clone has an overlap with p0439B06 clone (DBJ : AP002882) at the position 1 to 45,130 of 5' end. The sequence of this clone starts at the position 90,862 of p0439B06. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rtp.dna.affrc.go.jp/Genomeseq.html>.

Location/Qualifiers

```

1..126434
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="OSJNB0032H19"
/complement(3679..4206)
/gene="OSJNB0032H19.1"
/complement(3679..4206)
/gene="OSJNB0032H19.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB61171.1"
/db_xref="GI:14587253"
/translation="MARAILLHLSDPEVPLVYVAAYAAVALVLSCTSKHGKL
WKQOQRSGSSLPQIKAEESAGGAGSAROLIATISGIGRAAMAKVSNRRSR
OGCSSDDDRAGGGGLAVEEEDALMKRTITMGKCKPDPFSGHIAVDSGNLPRTTI
SKEANADAVANNIV"
/complement(join(4890..5008,6401..6539))
/gene="OSJNB0032H19.2"
/complement(join(4890..5008,6401..6539))
/gene="OSJNB0032H19.2"
/note="unknown protein"
/codon_start=1
/protein_id="BAB61172.1"
/db_xref="GI:14587254"
/translation="MQQVKRMDILPMLRDNQYHNGEFGVNLQNGSSVYTSSETL
IAEWRRLQVAVRVYVDLSVTSYSEKEDPILWIKFOR"
6879..7214
/gene="OSJNB0032H19.3"
6879..7214
/gene="OSJNB0032H19.3"
/note="contains EST AU082204(C52281)
unknown protein"
/codon_start=1
/protein_id="BAB61173.1"
/db_xref="GI:14587255"
/translation="MPSLRRLILAAAAVYVSLMPARVVADIDPSSGMPSPATTS
PFPCTTPACISTQFPVMAEPTTTTTFPDQDFGLASACHTAALAMIPVYFV
ESALVFLQ"
join(8942..8980,9207..9524)
/gene="OSJNB0032H19.4"
join(8942..8980,9207..9524)
/gene="OSJNB0032H19.4"
/note="contains ESTs AU101371(E20219), C99479(E20219)
unknown protein"
/codon_start=1
/protein_id="BAB61174.1"
/db_xref="GI:14587256"
/translation="MARWLQPCSLILRPTARRRTEAATRRRATRRPATRRR
RGSRRRRCRPTSRRCRRTPASGPTRRRAPRGARWCGRRRRWEPNCDRR
MVIYTAFLPFIKCS"
/complement(10569..10736)
/gene="OSJNB0032H19.5"
/complement(10569..10736)
/gene="OSJNB0032H19.5"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB61175.1"
/db_xref="GI:14587257"
/translation="MELKHGSIYDSNATTDRSIECDYIILDYLDMDAYVFMAMLI
DIDFNTLHSS"

```



```

107011 TTCAAAACATCGAGCCTCTTGCGACTACTGCTCT..... 107047
178 AlaIatrpSerSeraIaProGlyIleSerSeraIysSerSeraIeu 194
107048 .....GATCCTCGGATGATCGTACAAAGATGACACAGACTA 107086
195 SerSerThrIsgInGlyLeuAlaIatryGlyPheGlyLeuSerTrpIysAr 211
107087 TGCAG..... 107092
211 GAsnLeuLeuCyAsArgAsnCysAsp...ArgSerArgGlyGlyIuCYst 227
107093 .....CTATGCAGATGTGTGAATTCATGAGAAATGCTTCTCTGT 107135
227 hrIleGluGlu.....AsnSerSparGalaIatrr 236
107136 TCGTTCTCGAGTGTGAATCCCAAGTCATCTCCAGTCATCAAGCCTG 107185
237 CyAsArgTyrTrpCysIysGluAspIleHlsValSerIysLeuThrPhear 253
107186 TGCAGAGATACAAATT.....CCAGTACGTTTAGCTCTGATCTTCTC 107229
253 gCySlysValGluIutYtyr.....SerValItyrValLeuphe 265
107230 TTGCTGTGTGATCATCTTATAAAATATATATATCTCTT 107272
seq_name: gb_pl:AP003219
seq_documentation_block:
LOCUS AP003219 126434 bp DNA linear PLN 03-JUL-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, BAC clone:OSJNB0032H19.
ACCESSION AP003219
VERSION AP003219.3 GI:14587252
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (cultivar: Nipponbare) DNA, clone:OSJNB0032H19.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 126434)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
clone:OSJNB0032H19
Published Only in Database (2001) In press
2 (bases 1 to 126434)
Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannonada
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@agr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Jul 3, 2001 this sequence version replaced gi:13620984.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as Slicepredictor
(October 1998 version). The genomic sequence was searched against
NCBI NonRedundant Protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from M13rev to -21M13 of the BAC
clone. This sequence of OSJNB0032H19 clone has an overlap with the
P0439B06 clone (DDBJ : AP002882) at the position 1 to 45,130 of 5'

```

FEATURES

end. The sequence of this clone starts at the position 90,862 of P0439B06. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>. Location/Qualifiers

| | |
|--------|---|
| source | /organism="Oryza sativa" |
| | /cultivar="Nipponbare" |
| | /db_xref="taxon:4530" |
| | /chromosome="1" |
| gene | /clone="OSJNB0032H19" |
| | /complement(3679..4206) |
| | /gene="OSJNB0032H19.1" |
| | /complement(3679..4206) |
| CDS | /gene="OSJNB0032H19.1" |
| | /note="hypothetical protein" |
| | /codon_start=1 |
| | /protein_id="BAB61171.1" |
| | /db_xref="GI:14587253" |
| | /translation="MARAILLHSDTEPLVIVVIAVAALVIVSLCTSSKHGKLWKQOQSRGSGSLPQLKAESESGGSGSRKOLLATLSIGSKAAAKVSNRRSSRCGSSDDDEPGCGGLAVEEBEALMRTTIMGDKCRPLDPSGHIAIDSDGNLPATTISKANADAVNNITY" |
| | /complement(join(4890..5008,6401..6539)) |
| | /gene="OSJNB0032H19.2" |
| | /complement(join(4890..5008,6401..6539)) |
| CDS | /gene="OSJNB0032H19.2" |
| | /note="unknown protein" |
| | /codon_start=1 |
| | /protein_id="BAB61172.1" |
| | /db_xref="GI:14587254" |
| | /translation="MOQVKKRMDILPMLRDNVINGEFOVNLQVNGSSVYTSSETLIAEMKRLQVAVRTVVDLISVYTSSEKEDPILMLIKQR" |
| | 6879..7214 |
| gene | /gene="OSJNB0032H19.3" |
| | 6879..7214 |
| CDS | /gene="OSJNB0032H19.3" |
| | /note="contains EST AU082204(C52281)" |
| | unknown protein" |
| | /codon_start=1 |
| | /protein_id="BAB61173.1" |
| | /db_xref="GI:14587255" |
| | /translation="MPSLRPEILLAAAVVYSLMPARVADIDPSSGMPSPRATSPFPQPIIPAGISTQSPFMAEPEFTITTFPPDQGLASACHARTALAMIPVVFISAFLEVLQ" |
| | join(8942..8980,9207..9524) |
| gene | /gene="OSJNB0032H19.4" |
| | join(8942..8980,9207..9524) |
| CDS | /gene="OSJNB0032H19.4" |
| | /note="contains ESTs AU101371(E20219), C99479(E20219)" |
| | unknown protein" |
| | /codon_start=1 |
| | /protein_id="BAB61174.1" |
| | /db_xref="GI:14587256" |
| | /translation="MARWLQPCSLLRPTARRRRTEAATRRRATRRPATRRRGGSRRTTCRPTSRSGSTRRTASGPTRRRAAPRGAWCSRRRWPEPCDRRMYINTATLIPFGS" |
| | complement(10569..10736) |
| gene | /gene="OSJNB0032H19.5" |
| | /complement(10569..10736) |
| CDS | /gene="OSJNB0032H19.5" |
| | /note="hypothetical protein" |
| | /codon_start=1 |
| | /protein_id="BAB61175.1" |
| | /db_xref="GI:14587257" |
| | /translation="WEKHSIVDSNAATTDRSIECDIVILDVLLDMDAYFMAVLDIIFTLHSS" |
| | 14940..15419 |
| gene | /gene="OSJNB0032H19.6" |
| | 14940..15419 |
| CDS | /gene="OSJNB0032H19.6" |
| | /note="contains EST C27795(C52872)" |
| | unknown protein" |

```
/codon_start=1
/protein_id="BAB61176.1"
/db_xref="GI:14587258"
/translation="MASYRLVILVFSALVALAAGTTPADCPYCLLEPPPTVTTDC
PPPTSPSSGSIPTPPSSSSNTPPSSSSITWNPPOGGGGGIPYQPPAGGGGGG
GFNPAPPNPPIVPMYPMYRSPSPSPATVAVRGRSLASAVATAAALITVF"
complement(20579..21709)
/gene="OSJNB0032H19.7"
/codon_start=1
/complement(20579..21709)
/gene="OSJNB0032H19.7"
/translation="MESPTAAAMAAERPEEAALRALVDNRVAGVEAREVRL
TRSSRRHRLAAVEPLVAMLRSPADGFAALLALAVRERNKTKIVDAGALE"
complement(31272..32073,34221..34489)
/gene="OSJNB0032H19.11"
/translation="MGLGNVAILVSGIIGSVLVGGDAKLPSAGEVLSGAKEFKVKKH
NEGDTSSNDTAHTAOLLSOVNHLROEISLGRPTVVTNARSGPPTTVAVA
GANGVAYIKKGMKLSIDMEYTRKGLSDANVYVSGQLDVSVDYTSARRHLAGRIDR
DISLDETOEIESTRDEYTVIHGDLAFQDLSVNLVVRSLKLVSEIVQVPAS
VPLAIGSELVVRVSSLPQSTALPLVPTTPPAESPRAEAFQEQWQFVSRITSSCR
EGSGRFQEQRSVYTRTSSIREGSPSSNGASSSTGASTGNTSTSRFAMNASSRATPW
SLMS"
complement(join(28476..29162,29268..29441))
/gene="OSJNB0032H19.10"
/complement(join(28476..29162,29268..29441))
/gene="OSJNB0032H19.10"
/codon_start=1
/product="putative pollen-specific membrane integral
protein"
/protein_id="BAB61180.1"
/db_xref="GI:14587262"
/translation="MTTDAHAKKVDVYVNGDGEHVGEAOAHDLHEEAAAAAADH
HATRGIAIGFLIREVYEGSLASFLVYRSGVALLDEMGKGLTFPPVACIVAMATFV
LSMLGPAHFNPAVITTAATREFPWEKDLVLAADLAGSLACLSVNAEMPRHDF
YGTAPVYVHGTRLPFLMEFLASAVLMLVIAVTDAGTACVGGALIGAAYVGLGVI
GPVSGSNPAPRTLGPALIVGRYDGVAVYVAVADGMLVCAICNRAVRLSHRIYVFLC
GTSVGIAGSP"
complement(join(31272..32073,34221..34489))
gene
```

```
CDS
/gene="OSJNB0032H19.11"
complement(join(31272..32073,34221..34489))
/gene="OSJNB0032H19.11"
/translation="MESPTAAAMAAERPEEAALRALVDNRVAGVEAREVRL
TRSSRRHRLAAVEPLVAMLRSPADGFAALLALAVRERNKTKIVDAGALE"
complement(31272..32073,34221..34489)
/gene="OSJNB0032H19.10"
/complement(join(28476..29162,29268..29441))
/gene="OSJNB0032H19.10"
/codon_start=1
/product="putative pollen-specific membrane integral
protein"
/protein_id="BAB61180.1"
/db_xref="GI:14587262"
/translation="MTTDAHAKKVDVYVNGDGEHVGEAOAHDLHEEAAAAAADH
HATRGIAIGFLIREVYEGSLASFLVYRSGVALLDEMGKGLTFPPVACIVAMATFV
LSMLGPAHFNPAVITTAATREFPWEKDLVLAADLAGSLACLSVNAEMPRHDF
YGTAPVYVHGTRLPFLMEFLASAVLMLVIAVTDAGTACVGGALIGAAYVGLGVI
GPVSGSNPAPRTLGPALIVGRYDGVAVYVAVADGMLVCAICNRAVRLSHRIYVFLC
GTSVGIAGSP"
complement(join(31272..32073,34221..34489))
gene
```

alignment_scores:

| Quality: | 188.50 | Length: | 199 |
|---------------------|--------|-------------------|--------|
| Ratio: | 1.597 | Gaps: | 10 |
| Percent Similarity: | 59.296 | Percent Identity: | 29.146 |

alignment_block:

US-09-805-919-1 x AP003219/rev ..

Align seg 1/1 to reverse of: AP003219 from: 1 to: 126434

```
6 LeuLeuPheAlaCysValLeuLeuLeuValIhrLeuIleSerSerSerLy 22
|||||
116447 CTCATCTTTGCTATCTGACCTTCTCGACAGAGTTGC..... 116409
22 sAlaGlnAspIleSerGlnCysValProSerSerCysGlyAspIle...G 38
|||
116408 ...CAGGACACAGCAGTGTGCCACCTTCTCCCGCCGACATCTCA 116363
38 InLeuSHePheProPheArgLeuArgThrAspProGlnHisCysGlyArg 54
|||||
116362 ACATAGTATCATCTTCTCGCGAGAGCGACCCCTCAAGTGGGGTTT 116313
55 ArgGlyTyrGluLeuAspCysGlnAsnGlnThrValPheAsnTyrLy 71
|||||
116312 TCATCATAGCTAGCTAGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG 116263
71 sSerArgIlePheAspValGlnGluIleAsnTyrArgSerTyrSerIleA 88
|||||
116262 TACAGACATCTATGTCATCTAGCTACATCACTACACATCTGCTTCT 116213
88 rGluLeuAspProGlyLeuAsnAspGlnArgGlnCysThrVal... 103
|||||
116212 GGTCTCGTACGCGCAACCTGACGATGCAACAGCACTGCCCTTCTCT 116163
104 PheProAsnHisArg.....AlaSerTyrAs 112
|||||
116162 CGCTCCAAATCAATGCTCTTTCATTGGGGGATTCAGGGCCAGCTACAC 116113
112 P...AlaMetThrSerGlnIlePheGluThrValArgValAsnAsnAspI 128
|||||
116112 CAGGCAATGACTCTTCAGGCAATATGCTGGGCG..... 116079
128 LeuAsnTyrValAsnCysLeuAlaProIle...AsnSerSerGlnTyrIle 143
|||||
116078 ..ACCTTGTCAATTGTTGCGACGAGCAAGAACAAATACAAAGACATA 116031
144 ProHisSerPheCysSerTyrAsnSerThrGlyPheSerTyrLeuVal... 159
|||||
116030 CATGTGGAT...TGCCAGAACAGAGCTCTTATTTGCTATGTGCTGAA 115984
160 .....IleArgGluIleLeuGlnAlaSerAspLeuA 170
|||||
115983 TGTCCGACACATATTACGCTCTATGGAATACTTACGCTCT.....T 115940
170 IeAlaGlyCysArgValGluThrValAlaIlePheSerSerAlaProGly 185
|||||
115939 GCGCAATCTGTCGCAATGATTCAGATTAGTGCACCTCCGACGTGT 115893
seq_name: gb_PL:AP003250
seq_documentation_block:
```

LOCUS AP003250 146670 bp DNA linear PLN 01-SEP-2001
 DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone: P0443D08.
 ACCESSION AP003250
 VERSION AP003250.3 GI:15408719
 KEYWORDS
 SOURCE
 ORGANISM
 Oryza sativa (cultivar: Nipponbare) DNA, clone: P0443D08.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE
 AUTHORS 1 (bases 1 to 146670)
 TITLE Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0443D08
 JOURNAL Published Only in Database (2001) In press
 REFERENCE 2 (bases 1 to 146670)
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Direct Submission
 JOURNAL Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
 Agricultural Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rtp.dna.affrc.go.jp/,
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 On Aug 31, 2001 this sequence version replaced gi:14522861.
 Genes were predicted from the integrated results of the following:
 GENSCAN1.0, BLAST2.0, BLASTX2.0 as well as SplicePredictor
 (October 1998 version). The genomic sequence was searched against
 NCBI Nonredundant Protein database, nr
 (ftp://ncbi.nlm.nih.gov/blast/d) and the cDNA sequence database at
 RGP. Protein homologies of the coding regions were searched against
 NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent
 the identified cDNA sequences using BLASTN 2.0 with the
 corresponding DBJ accession no. and RGP clone ID.
 A gene with identity or significant homology to a protein is
 classified based on the protein name to indicate the homology level
 such as same name, 'putative-' and '-like protein'. A gene without
 significant homology to any protein but with EST homology (covering
 almost the entire length of partial sequence) is classified as an
 'unknown' protein. A gene predicted with a gene prediction program
 is classified as a 'hypothetical' protein.
 The orientation of the sequence is from 5' to 3' of the PAC clone.
 This sequence of P0443D08 clone has an overlap with OSJNB0083M16
 (DBJ: AP003214) at the position 1 to 624 and with P0480E02 (DBJ:
 AP002913) clone at the position 141,467 to 146,670 of 3' end. The
 sequence of this clone starts at the position 138,088 of
 OSJNB0083M16 and ends at the position 5205 of P0480E02. Detailed
 information on overlap and assembly quality together with
 annotation of this entry is available at
 http://rtp.dna.affrc.go.jp/GenomeSeq.html.
 FEATURES
 Source
 1. 146670
 Location/Qualifiers
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0443D08"
 2727..3227
 /gene="P0443D08.1"
 2727..3227
 /note="contains ESTs D48568(S14863), AU089765(S14863)
 similar to 16.9 kDa heat shock protein
 unknown protein"
 /codon_start=1
 /protein_id="BAB64123.1"
 /db_xref="GI:15408720"
 /translation="MSLVKLFDFAPDANPESIFGTVAADAMLSADTSAPANTYIE
 SRTAEAVPRPADLPAGVKKEVEVEGNVLTIGERSVREERKGGOSHIERSCA
 TEFGRHLPDADYDLVRAASMDGMLTVTPKVTYDKQRAIAAAVPRVAPVPAK
 ALEASP"
 complement(join(4071..4126,4283..4398,5525..5601))
 /gene="P0443D08.2"
 complement(join(4071..4126,4283..4398,5525..5601))
 CDS

gene
 /gene="P0443D08.2"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB64124.1"
 /db_xref="GI:15408721"
 /translation="MGVAMFIKINPKFDPNAKFFPRVCAPOSTACKACKLVRHHGEY
 AGIQSVNRPSPMLSLPFSFGVNGLLSATYRRILYTH"
 complement(6051..6500)
 /gene="P0443D08.3"
 complement(6051..6500)
 /note="contains EST AU093701(C63333)"
 /codon_start=1
 /product="putative 16.9 kDa heat shock protein"
 /protein_id="BAB64125.1"
 /db_xref="GI:15408722"
 /translation="MSLVRSNVDFDFADFPDFGVFSRLVTPATSDRTAFAANAY
 DMKETPESHVEKADLPGVKKEVEVEGNVLTISGSRSEKEDKNDKMRVERSSG
 QPMRRFLPBNAPVDVKAEMENGVLTVPAKAVKKPEVKAIEISG"
 complement(8660..10035)
 /gene="P0443D08.4"
 /note="pseudogene
 similar to non-LTR retroelement reverse transcriptase"
 /pseudo
 complement(8660..10035)
 /gene="P0443D08.4"
 complement(10583..11035)
 /gene="P0443D08.5"
 complement(10583..11035)
 /gene="P0443D08.5"
 /note="contains EST AU092880(C60687)"
 /codon_start=1
 /product="putative 16.9 kDa heat shock protein"
 /protein_id="BAB64126.1"
 /db_xref="GI:15408723"
 /translation="MSLVRSNVDFDFSLDMDPDSVRSVPATSDNDTAFAANAR
 IDMKETPESHVEKADLPGVKKEVEVEGNVLTISGSRSEKEDKNDKMRVERSS
 GQFMRRFLPBNAPVDVKAEMENGVLTVPAKAVKKPEVKAIEISG"
 complement(10583..11035)
 /gene="P0443D08.6"
 complement(10583..11035)
 /note="contains EST AU092880(C60687)"
 /codon_start=1
 /product="putative 16.9 kDa heat shock protein"
 /protein_id="BAB64127.1"
 /db_xref="GI:15408724"
 /translation="MSLVRSNVDFDFSLDMDPDSVRSVPATSDNDTAFAANAR
 IDMKETPESHVEKADLPGVKKEVEVEGNVLTISGSRSEKEDKNDKMRVERSS
 GQFMRRFLPBNAPVDVKAEMENGVLTVPAKAVKKPEVKAIEISG"
 complement(10583..11035)
 /gene="P0443D08.7"
 complement(10583..11035)
 /note="contains EST AU092880(C60687)"
 /codon_start=1
 /product="putative 16.9 kDa heat shock protein"
 /protein_id="BAB64128.1"
 /db_xref="GI:15408725"
 /translation="MEGSHGCGWNTASHFVDEREKRPVKSFWLVLRLLCCRRK
 SYRDIYAAGEEDGSEKSSLLTSSRLSLVSDADDAIDDAVYCRNASCAR
 KDQCAPVYVLPPIGHLHPVARAHGMVTTSSCGSGDGAAGCRRRFRMGGLRRLLM
 RRPMPVYVLPPIG"
 complement(join(19682..20619,20791..21045,21998..22714,
 23306..23926,24506..25205))
 /gene="P0443D08.8"
 complement(join(19682..20619,20791..21045,21998..22714,
 23306..23926,24506..25205))
 /note="contains ESTs AU030772(E60216), AU030771(E60216)
 similar to wall-associated kinase
 unknown protein"
 /codon_start=1
 /protein_id="BAB64129.1"

```

/db_xref="GI:15408726"
/translation="MSPSEFFVYVWSAMSLIMLMAARAGAESEGGGGLGKXCGDLN
ISSPFIIOGADRCGGPILDYVYCNNSNGVATIRSESTGSDIINISGDPMTVDL
VHKIARLNSTGCSIPYENTFAKLPITFTTISRNHLYVTNTEAPPAQOQQLVE
TRCENNTFFARLGRFHEGSDYDKYLEGGSRNSTYFLPLEPDPDGAASRYVELVG
GFLITMDLPAPVYSSAANTAADNRPOBGCAATVCCAKVISPFAVPAQAEKSCGM
LGFEVICHNDTPYLYGYKPRYRIQIIDIFYGNNSLVSDIHLGDFIYSGYKSCYC
HVPRTNKKVGLPEFSISTNTNLFLYSCNKAFLPDGSDDLVETRCGAKTRARGN
YSDGDPATFYMGCAATVYVLTGADARSYEOLIRGFFLLTQENPSSAFLELPA
SSSPGCLPTPCCKRTISYVPLPEPGRPCGSPQWQLTGMATLTITTEARVY
DITGNITVAVDENLPLATGCPAPPNISDGTQAFYISSNKAELRLSCNKPLPA
AAAPGFHSLPCDDQNSVRLVSDHHLHEDGLPPGCFVYVPIVORHNSMAGYIASM
RSGFLLEMAVSDGDCQVSGCCTSDLDLEACQPDGMHDKREFRKESEERAG
SSGGILIVSIFIFAMHKKRKQTDLDLKMSSSSQSGKDLGSPHIFTEE
LEBTAIGFASARELDGEGFTYKRGKLDGQRYVAVRYLYKNRYRVEGLNEVDLIS
LHONLITVIGTSSRSRDLIYEVYIPIGTAVYDLHGPBAGEGRTIMPEYRTIAET
AEALAYIAVEITHRDYKTNILIDNNHFKVADSLPSLELYETVSTVPGTGY
VDPYHOCYKLTDKSDYVSRGVYLIELISKPAVDSRSHSDINANMLNTOHBEV
DQIVDPGIEYDETRKMYDLVLAELAFQLOLDRESRPIKEVEVVLNCITNGECPA
EKMKNKASPKEDSHLKDLSLYSPDSVIRHPSQSTNHSVANSNG"
/gene="P044308.9"
/complement(join(27648..27664,28123..28903))
/gene="P044308.9"
/complement(join(27648..27664,28123..28903))
/gene="P044308.9"
/feature="hypothetical protein"
/codon_start=1
/protein_id="BAB64130.1"
/db_xref="GI:15408727"
/translation="MHPTLLCLPLASLLILCHRAACEPATGNTLVYFVWLGSP
NFNOSNOSSSALASCGHPAFVWNGVAYSLGSOILVLSIDYSSSEFAHARVA
DGGGVGRTDPTNSSLSPITSSNNAICFLSCNTEPEPIDGLVNAITSSCK
PIYAVIGIYIDRNPAPKAKNGCTSYVPLPMDSRPAITACTNYSPOKRFVLEMO
KNGRDCDACGSGGCGRTINDSAAAPACTCSGKRLRSTCGQNSF"
/complement(join(29901..29938,29998..30738,30790..30937,
30978..31158,31206..31362,31389..31517,31689..31800,
31910..32176))
/gene="P044308.10"
/complement(join(29901..29938,29998..30738,30790..30937,
30978..31158,31206..31362,31389..31517,31689..31800,
31910..32176))
/gene="P044308.10"
/feature="hypothetical protein"
/codon_start=1
/protein_id="BAB64131.1"
/db_xref="GI:15408728"
/translation="MGRSRPPOGROHNDKDTKAMPRESRHOGCBRRPKEMTSR
ENSPREKGIASAKETTPKSYAASPRNHOKRLSPULSTLESCRTPEGMPPRP
SRHPASLPALABEGEIGRRRVGPTSRHAPNATTAPEQOTAVRPMRRADHG
PRSGRTLEGRKRWQLAPSRRAAAAAPTPGRAPVDHVPRAAPASRDTRAASRP
SRGPRRAATPRLPATPRGRRAAGEWPSRADPDVGRPDLAKEVDPDPPRAAPS
SPPAVERVATVEGIAAAGGVKKEGEGWRLRFPSPRSREDAQDLFEFLVLY
LLAVFATNAHOLYGESCAPACAGNLTIKYPWLRGROPAYCGHPTFAVTCDDDDPT
GATASPPSINSYLRVLAHYGNSSVAFHNVLSACRATREMSSLAISLAV
SRANAGLPSANCRTPTGSLPNCVIGPSGGSEWFLSNRMTDQGRPARAYDVGCC

```

alignment_scores:

| Quality: | 174.00 | Length: | 286 |
|---------------------|--------|-------------------|--------|
| Ratio: | 1.137 | Gaps: | 14 |
| Percent Similarity: | 53.497 | Percent Identity: | 25.524 |

alignment_block:

US-09-805-919-1 x AP003250/rev ..

Align seg 1/1 to reverse of: AP003250 from: 1 to: 146670

```

6 LeuLeuphealAcysValLeuLeuValThrLeu...IleSerSerse 21
   : : : : : : : : : : : : : : : : : : : : : : : : :
36955 ATGATCTGTTAGCATGCTACGCTGATCTGCTCCCATGCTAGCTC 36906
   21 rLysAlaGlnAspIleSerGln.....CysValp 31
   : : : : : : : : : : : : : : : : : : : : : : : : :
36905 CACCGTTGATGCATTAACACGACCAACCATCGCGCTTCTCGGAGC 36856

```

```

31 roSerSerCysGlyAspIleGlnIleLeuLysPheProPheArgLeuArg... 46
   : : : : : : : : : : : : : : : : : : : : : : : : :
36855 CCGGACGCGCGGGAACCTTACCATCGGCTATTCATTTCTGGCTCGCCGC 36806
   47 ThrAspProGlnHisCysGlyArgArgGlyTyrGlnLeuAspCysGlnAs 63
   : : : : : : : : : : : : : : : : : : : : : : : : :
36805 AACGATCCCGCCGAGTGTGGTACAGCACTTTTCAGGACAGCTCGACCA 36756
   63 nAsnGlnThrValPheAsnTyrLysSerArgIlePheAspArgIleGlnL 80
   : : : : : : : : : : : : : : : : : : : : : : : : :
36755 TGAATAGCTTCCCTCCCAAGACGATCGACGACGACCAATTCAGCGCA 36706
   80 lAsnTyrArgSerTyrSerIleArgLeuLeuAspProGlyLeuAsnAsp 96
   : : : : : : : : : : : : : : : : : : : : : : : : :
36705 TCTTACACACAACAGCTCTTCATGACGACCAACAGACAGCTGACGAC 36656
   97 GlnArgGluAsnCysThrValPheProAsnHisArgAlaSerTyrAspAl 113
   : : : : : : : : : : : : : : : : : : : : : : : : :
36655 GGCCAG.....TGCCTCATCGAGAGTTTGTCAACGCTCTCCGACCT 36612
   113 aMetThrSerGlnIlePheGlnTyrValArgValAsnAsnAspIleAsn 130
   : : : : : : : : : : : : : : : : : : : : : : : : :
36611 CGGCTCAGCGAG...TTCAGATTAAGCCCATTAACCGGAGCTCGCT 36565
   130 yVal.....AsnCysLeuAlaProIleAsnSerSerGlnTyrIleProThr 145
   : : : : : : : : : : : : : : : : : : : : : : : : :
36564 TCTCTACACATGC.....AGCCAGAGTCAGCGGACCTTCCATC 36524
   146 SerPhe.....CysSerLysAsnSerThrGlyPheSerTyrLe 158
   : : : : : : : : : : : : : : : : : : : : : : : : :
36523 TCTTGCGCCCGCGTAGTGTGCGAANAATGATGCTCCATGCTGATGC 36474
   158 vValIleArgGluIleLeuGlnAlaSerAsp.....LeuAlaG 171
   : : : : : : : : : : : : : : : : : : : : : : : : :
36473 CTGGCTGCGGGGGAAGTACMAAGCTGATGACGATTCAGACMACTGCCG 36424
   171 lYlGlyCysArgValGluThrValAlaIlePheSerAlaProGlyIleSer 187
   : : : : : : : : : : : : : : : : : : : : : : : : :
36423 GGAACGCAAGGCTGATGATCCCGGTGCTGGGTACGAGGCTGGCGTG 36374
   188 SerAsnLysSerSerThrLeuSerSerThrHisGlnGlyLeuAlaTyrG 204
   : : : : : : : : : : : : : : : : : : : : : : : : :
36373 GCCAAMAACCTACAGCGGCTC.....ATCAAGGGGTG 36342
   204 yPheGluLeuSerTyrLysArgAsnLeu.....LeuCysArgAsnC 218
   : : : : : : : : : : : : : : : : : : : : : : : : :
36341 ATTCTTCTCGACTACAGCGCGCGCGCGCGCGGCGGACATTCGCAAGTT 36292
   218 yAspArgSerArgGlyLysGlyLysThrIleGlnGluAsnSerAspArg 234
   : : : : : : : : : : : : : : : : : : : : : : : : :
36291 GCTCGAGGAGC...GGCGCTGTGTGCGGTAATGTCACCTACGATGGG 36245
   235 AlaThrCysArgTyrTyrCysLysGluAspIleHisValSerLysLeuTh 251
   : : : : : : : : : : : : : : : : : : : : : : : : :
36244 CTAGAGTGCACA.....TCCCCGAAAGGCGCTGACGCTCTCCGATTCAC 36201
   251 rPheArgCysLysValGluTyrTyr.....SerValTyrValL 264
   : : : : : : : : : : : : : : : : : : : : : : : : :
36200 GTGTGTAGTTCGTTCCGCTTACGATCCCATTCCTTCCCGTCCGTT 36151
   264 euphePhe 266
   : : : : : : : : : : : : : : : : : : : : : : : : :
36150 TTTTITTT 36143

```

seq_name: gb_pl:AP002747

seq_documentation_block:

LOCUS AP002747

DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone: P0698603.

ACCESSION AP002747

VERSION AP002747.2 GI:11071975

KEYWORDS

SOURCE Oryza sativa (cultivar: Nipponbare) DNA, clone: P0698603.


```
ORGANISM      Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
REFERENCE
AUTHORS      1 (bases 1 to 147857)
TITLE        Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GM3) genomic DNA, chromosome 1, PAC
clone:P0698G03
Published Only in Database (2000) In press
2 (bases 1 to 147857)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (07-JUN-2000) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@agr.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Nov 1, 2000 this sequence version replaced gi:9049480.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as splicepredictor
(October 1998 version). The genomic sequence was searched against
NCBI NonRedundant Protein database, nr
(http://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologues of the coding regions were searched against
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBI accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from 5' to 3' of the PAC clone.
This sequence of P0698G03 clone has an overlap with P0463F06 (DBI:
AP002867) clone at the position 1 to 57,093 of 5' end and with
P0494A10 (DBI: AP002541) at the position 97,778 to 147,857 of 3'
end. The sequence of this clone starts at the position 87,230 of
P0463F06 and ends at the position 50,080 of P0494A10. Detailed
information on overlap and assembly quality together with
annotation of this entry is available at
http://rtp.dna.affrc.go.jp/Genomeseq.html.
Location/Qualifiers
1. 147857
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0698G03"
join(411..1189,1994..2088,2436..3508)
/gene="P0698G03.1"
join(411..1189,1994..2088,2436..3508)
/gene="P0698G03.1"
/codon_start=1
/product="putative receptor kinase"
/protein_id="BAB17321.1"
/db_xref="gi:11071976"
/translation="MDFTNLLIYVLLIAPIKYSDVATASDDEDFKTCSSHRCSKH
GPEIRPERLSTOPSCGAPGKOLSCSDQILDPVYSGCKVMYIRVHMAIPL
VDSRRLHLKPIISYNOSTAVRPTLKAASVLCSSIDPTQNSIVPTSLIPLA
NNASOFWLYAYEYMSILPGCTIVSDIPMPYSYDKNGPNFDSIFETAKAVIST
GTFVFWSSNVSTSCOCGHRGRCGSORADPOHSHMVEPTRVPESTLIRLT
CWCIFPAMICKIEELVSSVGTFTLSLIVATALYISLKSRYNEHLIKEMELKTYG
TSKPTRTISEVKTARFRKDLGAGAGVYKGGELNGVFAVKMLENSVGBSEFT
NEVATIGRIHNAIVRLGFCSEGRQALIEFMPNSLEKTIIPHSNSIRRELVAD
KMIDLALGARMEYLHQCNCORILHPIKPHNILLDSYFNPKISDGLAKLAROS
IYTLAARGTKVYIAPELVSRTFGAISYKSDVYSGLVLEMSGRNTPETLONOME
FPEPEIYRVMAGODLVLTMETTQGEKEMVROLAIVALMICQWPNKRPDSMTKVVM
LGRILNLOVPPKPTSSONOLVT"
complement(join(4690..4841,5041..5231,5395..5432))
/gene="P0698G03.2"
complement(join(4690..4841,5041..5231,5395..5432))
FEATURES
source
1. 147857
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0698G03"
join(411..1189,1994..2088,2436..3508)
/gene="P0698G03.1"
join(411..1189,1994..2088,2436..3508)
/gene="P0698G03.1"
/codon_start=1
/product="putative receptor kinase"
/protein_id="BAB17321.1"
/db_xref="gi:11071976"
/translation="MDFTNLLIYVLLIAPIKYSDVATASDDEDFKTCSSHRCSKH
GPEIRPERLSTOPSCGAPGKOLSCSDQILDPVYSGCKVMYIRVHMAIPL
VDSRRLHLKPIISYNOSTAVRPTLKAASVLCSSIDPTQNSIVPTSLIPLA
NNASOFWLYAYEYMSILPGCTIVSDIPMPYSYDKNGPNFDSIFETAKAVIST
GTFVFWSSNVSTSCOCGHRGRCGSORADPOHSHMVEPTRVPESTLIRLT
CWCIFPAMICKIEELVSSVGTFTLSLIVATALYISLKSRYNEHLIKEMELKTYG
TSKPTRTISEVKTARFRKDLGAGAGVYKGGELNGVFAVKMLENSVGBSEFT
NEVATIGRIHNAIVRLGFCSEGRQALIEFMPNSLEKTIIPHSNSIRRELVAD
KMIDLALGARMEYLHQCNCORILHPIKPHNILLDSYFNPKISDGLAKLAROS
IYTLAARGTKVYIAPELVSRTFGAISYKSDVYSGLVLEMSGRNTPETLONOME
FPEPEIYRVMAGODLVLTMETTQGEKEMVROLAIVALMICQWPNKRPDSMTKVVM
LGRILNLOVPPKPTSSONOLVT"
complement(join(4690..4841,5041..5231,5395..5432))
/gene="P0698G03.2"
complement(join(4690..4841,5041..5231,5395..5432))
gene
/gene="P0698G03.2"
/feature="hypothetical protein"
/codon_start=1
/protein_id="BAB17321.1"
/db_xref="gi:11071976"
/translation="MCTPKLTQVHRLRTPEDRAVAAYVIGSLPSSPMARLGAIV
RLSGSARRRSGEGCGRRNGERENCAAEASATRGITAVAAEHGHSVRRQRP
EGEERGEKKRREKKRREKKRMS"
complement(join(926..8061,8244..8395,8496..9262))
/gene="P0698G03.3"
complement(join(926..8061,8244..8395,8496..9262))
/gene="P0698G03.3"
/codon_start=1
/product="putative receptor kinase"
/protein_id="BAB17321.1"
/db_xref="gi:11071978"
/translation="MWIPGFFRSALQVFSVLCVLFVPPDIDEMPLHIGIPAC
PFTGSLHSLASVPPHRRGVEGVSAYELCTDDKATIQIDNGTYEVTGIDYDSFTFW
VDANTSDRRNCPLPRMKRIPIYVDVSDSDSSPNIDQVLELDPASRMSPFNCSX
EINNVMSPVACMGTSSTFYVITGLSCYINTEPSCGILAMPIDMGVAPAYD
SLSIDVYKFRKFRVQFPFRHRRNSFKDCIMESIPKSTIGIKISDNFSVTEFN
MDCATKIPSPYNILGLILFALMFMTLAPPCRYLIAPLVILFIAQYFLKRMITID
AVERFLRMQMLGPTRAYVDIILAIFHFKELGQGVSVYGVLLPDDLHVALKITL
EYNSCNGEERISEVATIGRIHNVNRYLVGFCSEKMRALVEYMRPSGLKHTFS
RRSEMDKLNEIALGAGIYHVGQDMQIILFDIKPHILDDNPEVADFGIAK
LYPRKSPVSRALRGTYGVAPAPMSYRSFGIISGSDVYSFGMLLEWYGRNADP
NADSSAKAYYPSWVYDKLIDQVDEISNPANMHELEKICVIGLQICQMSHDEPT
MSEATHEMLEGVDALQVLPFRFFCDGMPPOVMDSYFHSLSLASEEDDGLAELA
join(9938..10001,10770..11578)
/gene="P0698G03.4"
join(9938..10001,10770..11578)
/gene="P0698G03.4"
/feature="hypothetical protein"
/codon_start=1
/protein_id="BAB17321.1"
/db_xref="gi:11071979"
/translation="MTCNVYISLSYFLNYSKFTPPPPRIPHPAPARLNEGII
PDSLSRPFRLSLSLGFVIAIADIVNGARELRGFPSPAPASFRRKATVP
RASLFRFHSPLALAVGVPAPCSLSAELFAVAVPVLPPRHLHLRLRYTA
DVPKRSAPSLARSRSTVIVINRAAFLRSGRRSRRCGVSRRPPAPLISALSG
AAPPRMSPATIGARPPVVGAGRAKALAAKAKRPSAARSRVYDAADAM
APTVPYPRASLIG"
join(14575..14697,15133..15179,17078..17656,17671..18037,
18110..18307)
/gene="P0698G03.5"
join(14575..14697,15133..15179,17078..17656,17671..18037,
18110..18307)
/gene="P0698G03.5"
/feature="hypothetical protein"
/codon_start=1
/protein_id="BAB17321.1"
/db_xref="gi:11071980"
/translation="MESPTMPTGTFGVDGPEGEVSALLISPPGRVAPAPAPARAETA
TERRACPRKRSRSGSADVHKGWGRVSPYTHGGRAPVAPVPRVSRFG
RCRMTAGTTPGRRARRRRRRRRRARRARAPVTRGDRGGAPEAREEGSEER
NSAPAPGTTTTRTACVVEEGCAARDDDGATVTERKNGADENDDGAAKREYAP
SSEEDRSDGGEPEKRRKATAFKMPRAIVLHGADVGNATLAKMVSRSRSDSD
VGEERKVDGDEGRCARRHESDSEERRAAETERTGSKYATAGGEGREKRP
AEGISALIDAGELRGVNDDESGEGREGGTGIGTGGMWRGGRKMTGSHPS
ARVAGCPACQRRARAGEPMGGERREGELGRAGPREEGGEGF"
join(21557..22009,24616..25161,25209..25575,25648..25845)
/gene="P0698G03.6"
join(21557..22009,24616..25161,25209..25575,25648..25845)
/gene="P0698G03.6"
/feature="hypothetical protein"
/codon_start=1
/protein_id="BAB17321.1"
/db_xref="gi:11071981"
/translation="MMHGMRLIGCMRCADICRSGSRAADVHKGWGRVSPYTHGGRG
ASRRRSAPVAPVRSRTGRRLITAGPTGRRARRRRRRRRRARRARAPVTRG
HRRGAPAPERARERGESERRSSAPGTTTTRTACVVEEGCAARDDDGATVTER
NGADEVYGDAAAPREVAPESEEDRSDGGEPEKRRKATARGAADVGAGTAKMEVSP
```

```

SREGRSDVGEPEKVDGDDGGERGARGREHSDGSEERAAETEEENGSKYITAGGER
GGRERKRAEISGAIIDAGELRGVNDPDESGREKGGTGGITGGTCMRGRGKRA
MTGSHPSARVAGGACCOARARAGELMGGRERREBELGRAGPRDEGGEDF"
/gene="P069803.7"
/join(26540..28903,28967..32693)
/gene="P069803.7"
/join(26540..28903,28967..32693)
/codon_start=1
/product="putative polypeptide"
/protein_id="Bab1737.1"
/db_xref="GI:11071982"
/translation="MAEQAVHDLSVSGDGEPPRRRATPPPPHONKQGEAL
ERKESASVAGAGRDGERLRLVYDGSPPQALDAVALLRHPVDPDESPA
QRMIDVNLVMTAQRAGSRATKTSGATGTSASRRARAAVAHRSATP
SSAPTRDEHGEEDARDIERRNSRRTPRATEGSSRVSPROREDOPMPAG
VGCAPVSLRVNWRPRRTTEKVDGVSNTLETQYTTGIEAGGDDVMAFF
PMALGQARMLNLPASVHSMEDLCQOTFMNPGSTYRPEEDLHAYOGDESL
RSYTORCQYRNTPCIPAHAYVYARGGYRHRMLEKASKEPOTATPOLARVA
KKEAWTWNPSGGSVASAPGSAQGTGRDRKRKRSVHSDGEHVLVEGVPATR
KGRASDKKEAGTSPSERPASKVHNTSLDLDCRAVKNLERTKMEDEBROE
RREKSSAIVSGKRSEAKQKAPVADIDDDDLGQEBGATLAVDGGACAHVSHS
FKARRRELLAAPTHTATRRARMSVALLTDDHPGCVARGQIMVYSPITVKL
GRVLIDGALNLPAPAPAIKAPGVLOSPGIIIGVTHPLGHIIDLVTGGS
ANFTEEVNPDVADLSIPYAVYGRPAIKEMAVATLOKMKMGPGPISITRDLK
VALACMQRADHLAATKPPGDEGRLGTSAPTAPRRTITCEVPYKEDALVSTRA
MTDFARPRADMPGVPREVLEHRLAVRPAKRVKQVROQAPEROFLIEEVARLEA
CFIREVHPEMLANPVVPRANKRLMCIIDYDLNACRDEVPPLRDOIIDSTAGC
DLICFLAYAGYDQIRMARDEDEKTAFTPVGYCYTSSFGIKNAPGFOFTRTSL
GSOIGRVAEYVDLVTYKNOETLSDLAETENRSARKLNPKCYFGVPAKLL
GEIYASGIEANPEKIRAIEMRPPSKIRVOCVGMALSFRTISRGKALPLFKL
LKRSGPTTTEEAHEALITOKAYISSPVIYAEPPPELIIYATPOVYSKALTYER
DENRPHVPHASVITWPSERGEAPESDGLRLITGVPPLACQTVLAGDPDEGP
KATAGRPLHPDEANPLVRPKDGEAPENGGORPLTGVGPIACQTVLAGDPDEGP
DPOGPATGPRPLSSDEVTGIDKCAKPNLNEBRCGTAPSEKEDRPRRQVQRP
VYVESEALRDAKTRVPOAKMLYAILMASRKLHYQARVTVSVSPLGQILNREG
TGRVKNALIESPDHFEPRHAIKSOILADFAVETPSETYSIPEASTNSOLPH
TTHWMDGSLGLOGAGVTLTSPGDDVRYVLRDPFATNMAYEBGLLAGLVAA
GLIRRLIVGDSQIYAVOVCKEYRCSDPVYVNOVMEHRPHGJILREVRPN
MIADLSRLASSRAQTPPGAFERLTOPKARPPLETTAPDRPPPVVLASGPEGS
APSSRLIAMISETQYLTUTKLPEDGESERQISKRVYVEGLIYRANGCILK
CIPQEOVELLADIHEBEGCAHSRTLVGAKROGFYPTLVNADLVKRCACQF
HAKOIHOPALQIIPLSMPFAVWGLDILGFEFRAGGEYLVADIKFTKPEAVPY
VKIDKHSALKEIKGIFARFGVNRILINDGTSELGCEMDGKICFASPAHR
SGQERANAEILKLTKEITFNILKKGSDMIELVLMANETTPSRATGEPFLV
YGAEAVLPSELTLSPRATWYCEADODOLARDLDLLEERRRRAARARXOOSIARY
HORYRARSCLCVDIVLRVOTRAGLSKLSPMWEGYRIVGPRPSVYALAGDGTSL
PFWNIHLRRFTF"
complement(join(33311..34366,34450..35245,35345..36603))
gene
alignment_scores:
    quality: 167.50      length: 238
    ratio: 1.279        gaps: 10
Percent Similarity: 55.042   Percent Identity: 26.471
alignment_block:
US-09-805-919-1 x AP002747/rev ..
Align seg 1/1 to reverse of: AP002747 from: 1 to: 147857
2 leuThrArglyLeuLeuPheAlaCysValLeuLeuValThiLeuI 18
|||||
72773 TTGCAGACAGCTCTAGTGAATCTCTTTCTTTGCTGTGGACGAGATG 72724
18 eSerSerSerLysAlaGlnAspIleSerGlnCysValProSerSerCysG 35
|||||
72723 TGGAGGAGGAGGACAAAGACAAAC...CACTACTGCTCTCTCTCTCTCTG 72677
35 LysAspIle...GlnIleLysPheProPheArgLeuArgThiAspProL 50
|||||
72676 GTGATCTTCACAAACATATCCATCCCTTCGCTTCAAGCCATTCGGCG 72627
51 HisCys.....GlyArgGArgGlyTyrGlnLeuAspCysGlnAsn.. 63
|||||
SREGRSDVGEPEKVDGDDGGERGARGREHSDGSEERAAETEEENGSKYITAGGER
GGRERKRAEISGAIIDAGELRGVNDPDESGREKGGTGGITGGTCMRGRGKRA
MTGSHPSARVAGGACCOARARAGELMGGRERREBELGRAGPRDEGGEDF"
/gene="P069803.7"
/join(26540..28903,28967..32693)
/gene="P069803.7"
/join(26540..28903,28967..32693)
/codon_start=1
/product="putative polypeptide"
/protein_id="Bab1737.1"
/db_xref="GI:11071982"
/translation="MAEQAVHDLSVSGDGEPPRRRATPPPPHONKQGEAL
ERKESASVAGAGRDGERLRLVYDGSPPQALDAVALLRHPVDPDESPA
QRMIDVNLVMTAQRAGSRATKTSGATGTSASRRARAAVAHRSATP
SSAPTRDEHGEEDARDIERRNSRRTPRATEGSSRVSPROREDOPMPAG
VGCAPVSLRVNWRPRRTTEKVDGVSNTLETQYTTGIEAGGDDVMAFF
PMALGQARMLNLPASVHSMEDLCQOTFMNPGSTYRPEEDLHAYOGDESL
RSYTORCQYRNTPCIPAHAYVYARGGYRHRMLEKASKEPOTATPOLARVA
KKEAWTWNPSGGSVASAPGSAQGTGRDRKRKRSVHSDGEHVLVEGVPATR
KGRASDKKEAGTSPSERPASKVHNTSLDLDCRAVKNLERTKMEDEBROE
RREKSSAIVSGKRSEAKQKAPVADIDDDDLGQEBGATLAVDGGACAHVSHS
FKARRRELLAAPTHTATRRARMSVALLTDDHPGCVARGQIMVYSPITVKL
GRVLIDGALNLPAPAPAIKAPGVLOSPGIIIGVTHPLGHIIDLVTGGS
ANFTEEVNPDVADLSIPYAVYGRPAIKEMAVATLOKMKMGPGPISITRDLK
VALACMQRADHLAATKPPGDEGRLGTSAPTAPRRTITCEVPYKEDALVSTRA
MTDFARPRADMPGVPREVLEHRLAVRPAKRVKQVROQAPEROFLIEEVARLEA
CFIREVHPEMLANPVVPRANKRLMCIIDYDLNACRDEVPPLRDOIIDSTAGC
DLICFLAYAGYDQIRMARDEDEKTAFTPVGYCYTSSFGIKNAPGFOFTRTSL
GSOIGRVAEYVDLVTYKNOETLSDLAETENRSARKLNPKCYFGVPAKLL
GEIYASGIEANPEKIRAIEMRPPSKIRVOCVGMALSFRTISRGKALPLFKL
LKRSGPTTTEEAHEALITOKAYISSPVIYAEPPPELIIYATPOVYSKALTYER
DENRPHVPHASVITWPSERGEAPESDGLRLITGVPPLACQTVLAGDPDEGP
KATAGRPLHPDEANPLVRPKDGEAPENGGORPLTGVGPIACQTVLAGDPDEGP
DPOGPATGPRPLSSDEVTGIDKCAKPNLNEBRCGTAPSEKEDRPRRQVQRP
VYVESEALRDAKTRVPOAKMLYAILMASRKLHYQARVTVSVSPLGQILNREG
TGRVKNALIESPDHFEPRHAIKSOILADFAVETPSETYSIPEASTNSOLPH
TTHWMDGSLGLOGAGVTLTSPGDDVRYVLRDPFATNMAYEBGLLAGLVAA
GLIRRLIVGDSQIYAVOVCKEYRCSDPVYVNOVMEHRPHGJILREVRPN
MIADLSRLASSRAQTPPGAFERLTOPKARPPLETTAPDRPPPVVLASGPEGS
APSSRLIAMISETQYLTUTKLPEDGESERQISKRVYVEGLIYRANGCILK
CIPQEOVELLADIHEBEGCAHSRTLVGAKROGFYPTLVNADLVKRCACQF
HAKOIHOPALQIIPLSMPFAVWGLDILGFEFRAGGEYLVADIKFTKPEAVPY
VKIDKHSALKEIKGIFARFGVNRILINDGTSELGCEMDGKICFASPAHR
SGQERANAEILKLTKEITFNILKKGSDMIELVLMANETTPSRATGEPFLV
YGAEAVLPSELTLSPRATWYCEADODOLARDLDLLEERRRRAARARXOOSIARY
HORYRARSCLCVDIVLRVOTRAGLSKLSPMWEGYRIVGPRPSVYALAGDGTSL
PFWNIHLRRFTF"
complement(join(33311..34366,34450..35245,35345..36603))
seq_documentation_block:
LOCUS AF100765 3067 bp DNA linear PLN 02-AUG-1999
DEFINITION Oryza sativa receptor-like kinase (BARK1) gene, complete cds.
ACCESSION AF100765
VERSION AF100765.1 GI:5669662
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 3067)
Feuillet,C. and Keller,B.
High gene density is conserved at syntenic loci of small and large
grass genomes
Proc. Natl. Acad. Sci. U.S.A. 96 (14), 8265-8270 (1999)
99324224
MEDLINE
2 (bases 1 to 3067)
Feuillet,C. and Keller,B.
Direct submission
Submitted (23-OCT-1998) Plant Biology, University of Zurich,
Zollikerstr. 107, Zurich 8008, Switzerland
JOURNAL
FEATURES
1. 3067
location/Qualifiers
/organism="Oryza sativa"
72626 GATTGGTTCGTCGCTCCCGCTCATAGTATGACCTGCTTCGACCAACAC 72577
64 .AsnGlnThrValPheAsnTyrLysSerArgTllePheAspValGlnI 80
|||||
72576 TGGCAAGGCGCACCATTCAGATCAACACAAACAACTACTAGCTGAGTAGCA 72527
80 LAsnTyrArgSerTyrSerIleArgLeuLeuAspProGlyLeuAsnAsp 96
|||||
72526 TCACACACACCATCTGCTCTCTTCTTGCTGACCGACGACCATTCAGAGAT 72477
97 GlnArgGlnAsnCysThrValPheProAsnHisArgAlaSerTyr.... 111
|||||
72476 TGATGATAGAGCTGCGCTCTCTCTCCATCCATCCATCACCCTCTACAGTAC 72427
112 ..... 113
72426 TCAGCATCTGTTGGACCGCTACACAGATCTTATGCGCTTTCGATTT 72377
113 LAsnThrSerGlnIlePheGlnTyrValArgValAsnAsnAspIleAsn 129
|||||
72376 TGGCTACTCTTCAGACAGAGCTGCTGGCA.....TGC 72345
130 TyrValAsnCysLeuAlaProle.....AsnSerSerGlnTyrI 143
|||||
72344 TTGTGATATGTTCCAGACCACTACAGATCTACTACTGATGATGACAG 72295
143 eProThrSerPheCysSerLysAsnSerThrGlyPheSerTyrLeuValI 160
|||||
72294 GCCCGCTCT...TGCGGCCACCAACATCTTTGTTTATCTCAACG 72248
160 LArgGlnIleLeuGlnAlaSerAspLeuAlaGlyGlyCysArgValGln 176
|||||
72247 AGTTTNTGCTGTACAAATGTGAGAGCTTCAACCTCTTGTGCTACTGTG 72198
177 ThrValAlaTrpSerSerAlaProGlyIleSer.....SerAs 189
|||||
72197 TCATGATATCCCTTGATGATCTCAATATATACAGCTCACAGTACAAA 72148
189 nLysSerSerThr.LeuSerSerThrHisGlnGlyLeuAlaTyrGlyPhe 205
|||||
72147 TGCAGATTACTGATATCATTCATTCATGAGGAGGCTTGCTGTTG 72098
206 GlnLeuSerTrp 209
|||||
72097 GATTCACATGCG 72086
seq_name: gb_pl:AF100765
seq_documentation_block:
LOCUS AF100765 3067 bp DNA linear PLN 02-AUG-1999
DEFINITION Oryza sativa receptor-like kinase (BARK1) gene, complete cds.
ACCESSION AF100765
VERSION AF100765.1 GI:5669662
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 3067)
Feuillet,C. and Keller,B.
High gene density is conserved at syntenic loci of small and large
grass genomes
Proc. Natl. Acad. Sci. U.S.A. 96 (14), 8265-8270 (1999)
99324224
MEDLINE
2 (bases 1 to 3067)
Feuillet,C. and Keller,B.
Direct submission
Submitted (23-OCT-1998) Plant Biology, University of Zurich,
Zollikerstr. 107, Zurich 8008, Switzerland
JOURNAL
FEATURES
1. 3067
location/Qualifiers
/organism="Oryza sativa"

```

```

/db_xref="taxon:4530"
/Chromosome="1"
<331..1037
/gene="8ARK1"
join(331..1037,1153..1310,1386..2533)
/gene="8ARK1"
/note="similar to wheat ARK1AS"
/codon_start=1
/product="receptor-like kinase"
/protein_id="AAd6415.1"
/db_xref="GI:566963"
/translation="MAIHGEFALLILISAVISYLCILAVLPDPAAGRRHRRHRCPPPTC
GNLSVSPRRRGDPSECGVSLYELNCADKDKATIDNCTVYVGTINTDSELTLL
ADISDSPPSCPLPRWNLSRPERVMTFRASVYNSQELKNSMIMYFVACLSIR
SFVYLTGONPALIOSLEPSGGLIAMPPLGKROSAVAVNSLSLADVYKMRKGR
VQFPRRSFKQCPMDSIPTGNTMFGTILSADNPQLCIPIHQIPSRVMPGLIMLY
IAYFOIKMTAATCRFLPLPLAVFLAKWKRTYITDAVEKFLMOQMLPTRYA
YTDITAITGHRREKLGOGGYSYKGLPLGDLHAVKMSNGNSNGEIESEVATI
GMIHVNVYRLVGFCSSEMRRLAYEMPRGSLDKYIYLSERFSMDKINELALGTAR
GINLHOGCDMOITLHPDIKPHNLTIDNTPYKADNGIATKYPRDSFVPRALRGTV
GYMAPKMSRSGYISKSDVYSFGMLLIMAGRRVDPNADSKSKAYRPSWVDO
LIADQDELSIVANMHELEKRLCLVGLWCTOKSHDRPTMSAILEMREGVDALQVP
PRPFCDGDGIGNMPPQVMDSYFHSKLTALISEBDGIAKLASS"
<331..>2533
/gene="8ARK1"
join(<331..1037,1153..1310,1386..>2533)
/gene="8ARK1"
/product="receptor-like kinase"
1153..1310
/gene="8ARK1"
1386..>2533
/gene="8ARK1"

BASE COUNT      819 a      707 g      891 t      2 others
ORIGIN
1 MetLeuThrArgGlyLeuLeuPheAlaCysValLeuLeuValThrLe 17
   :::::
   :::::
   :::::
361 CTACTGCTACTCGGTAATTTCTGTCTGCTGTTCTTCTGCGATTCTTCT 410
   :::::
   :::::
   :::::
17 uileSerSerSerLysAlaGlnAspLieserGlnCysValProSerSerC 34
   :::::
   :::::
   :::::
411 CCCAGATGACAGTGGCGGCATCATCGCATCGCTCCCTCCGTTACCT 460
   :::::
   :::::
   :::::
34 ysgIyAspIle..GlnIleYsPheProPheArgLysArgLysPro 49
   :::::
   :::::
   :::::
461 GTGGTAATCTTAGCAGCGGTGCTGCTCTTCCGCGCGGAGAGTACCCA 510
   :::::
   :::::
   :::::
50 GlnHisCysGlyArgArgGlyTyrGluLeuAspCysGlnAsnAsnGlnTh 66
   :::::
   :::::
   :::::
511 TCTGAGTGGGAGTCTGTGCTAGAGCTCAATTGTGCCGATGACAGACGC 560
   :::::
   :::::
   :::::
66 rValPheAsnTyrLysSerArgLLePheAspValGlnGluIleSerTyr 83
   :::::
   :::::
   :::::
561 TACATTCAGATTGACAAACGAAACATCATGTGACGGGCACTCAACTACA 610
   :::::
   :::::
   :::::
83 rGserTyrSerIleArgLeuLeuAspProGlyLeuAsnAsnGlnArgGly 99
   :::::
   :::::
   :::::
611 CTGATTTCTACTCTTCTGATGATGATGACATCTCGGATTCACCTATAC 660
   :::::
   :::::
   :::::
100 AsnCysThrValPheProAsnHisArgAlaSerTyrAspAlaMetHis 116
   :::::
   :::::
   :::::
661 AGTGCCTCTT.....CCTGATGGAATTCCTTCGAGA 695

```

```

116 rGlnIlePheGlu.....TrrValArgValAsnAsnAspIleAsnTyrV 131
   :::::
   :::::
   :::::
696 GCCCAGATTCACAGTCTCTATGACACTGATA.....GCTACTCTTG 736
   :::::
   :::::
   :::::
131 alAsnCysLeuAlaProIleAsnSerSer.....GlnTyrIlePro 144
   :::::
   :::::
   :::::
737 TCAATGTTCGAGAAATPAAGAAATAGTAGCTGATTATGACTAGCT 786
   :::::
   :::::
   :::::
145 ThrSerPheCysSerLysAsnSerThrGlyPheSerTyrLeuValIleAr 161
   :::::
   :::::
   :::::
787 GTGGCT...TGCGTAGACACAGACATCTCGTTATGCTTACACGG 833
   :::::
   :::::
   :::::
161 gGlu.....IleLeuGlnAlaSerAspLeuAlaGlyGlyCysArgV 175
   :::::
   :::::
   :::::
834 CCAAAATCTCGTCTTATTCAGAGCTTGAGCTTCTGTGGTACTAG 883
   :::::
   :::::
   :::::
175 alGluThrValAlaIlePhe.....SerSerAlPro.....Gly 185
   :::::
   :::::
   :::::
884 CCAATGACTCTTGGGTGGTGGGACAGTGTGACACCACTAATGAAAGC 933
   :::::
   :::::
   :::::
186 IleSerSerAsnLysSerSerThrLeuSerSerThrHisGlnGly..... 200
   :::::
   :::::
   :::::
934 TTAAGCTTAAGCTATGACAGACGTAATATCTATGAGGAAGATCAG 983
   :::::
   :::::
   :::::
201 .LeuAlaTyrGlyPheGluLeuSerTrrPysArg 211
   :::::
   :::::
   :::::
984 AGTCAATTTCCTTCAGAAATAGCTTTAAGCAA 1017
   :::::
   :::::
   :::::

seq_name: gb_pl:AF238475
seq_documentation block:
LOCUS AF238475 3999 bp DNA linear PLN 21-JUN-2000
DEFINITION Oryza sativa receptor-like kinase (RLG17) gene, complete cds.
ACCESSION AF238475
VERSION AF238475.1 GI:8575483
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 3999)
Yun,C.-H., Lee,G.-R. and Kim,H.-I.
Molecular analysis of receptor-like kinase in rice
JOURNAL Unpublished
AUTHORS Yun,C.-H., Lee,G.-R. and Kim,H.-I.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-2000) CytoGenetics, National Institute of
Agricultural and Technology, Seodundong 249, Suwon 441-707, Korea
FEATURES
source
1..3999
/organism="Oryza sativa"
/variant="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
join(<787..1502,1673..1803,1915..>3020)
/gene="RLG17"
/product="receptor-like kinase"
<787..>3020
/gene="RLG17"
join(787..1502,1673..1803,1915..3020)
/gene="RLG17"
/note="Protein kinase"
/codon_start=1
/product="receptor-like kinase"
/protein_id="AAF78019.1"
/db_xref="GI:8575484"
/translation="MRVITYVLCVGVLPDPAAGRRHRRHRCPPPTCGHLSDFEP
RRGDPECGVQSYELTICADKATITQDKETYSVDINTGDSITLVVDAFSDRSRSC
LIPRNPLLRDPRLOAKSHHILIEAPVGVTAASFVCSQELNNSMIMYFVACLSIR
SFVYLTGONPALIOSLEPSGGLIAMPPLGKROSAVAVNSLSLADVYKMRKGR"

```



```

7556 AGGACATCATCTATCTGATCTGCTGGCTGCATGCCAGCTT 7507
94 uasnasplnargluasnscysthrvalheproasn..... 106
7506 CTGGATGTCACGTAGCAGTTGC...CTCTCTCTCTGCGGAACCCCTTC 7460
107 ..HisargalaserTyrAspAlaMetHSerGlnIleheglu..... 120
7459 TTGGTATGCCAGCTCCAGCTCCAGCACTCACACATCTTATATTGGCA 7410
121 .....TTPyAlaGValAsnAsnAspIleasnTyrVa 131
7409 CTTCCAGTAGGGGTACTGCTGCT.....ACGTTTCT 7378
131 lAsnGlyLeuAlaProIleAsnSer.....GlnTyrIleProIlns 146
7377 CAATTGTTCCGACGAAATPAGAACGACGCTGGATGATCATGCGCGTGG 7328
146 erPhecysSerLysAsnSerThrGlyPheSerTyrIleValIle...Arg 161
7327 CT...TGCCTGACACACAGATCCTTCTGTTTACTGTTTACTGGCCA 7281
162 GluIleuGlInalaserAspLeuAlaGlyGlyCysArgValGluThrVa 178
7280 CAATCTGCTTATATTCAGAACCTTGACCTCTGCTGGGTAC.....CT 7237
178 lAlaTrpSerSerAlaProGlyIleSerAsnLysSerSerThrLeus 195
7236 GGGCAGACACTCTTTGGTGGCAGCAAGTTGAACACAGCTGCGCGCTG. 7188
195 erSerThrHisGlnGlyLeuAlaTyr 203
7187 .....CAGAAATGTAAGCTAT 7173

seq_name: gb_pl:AC006418

seq_documentation_block: 67730 bp DNA linear PLN 05-APR-2000
LOCUS AC006418
DEFINITION Arabidopsis thaliana chromosome II section 247 of 255 of the
ACCESSION AC006418 AE002093
VERSION AC006418.3 GI:6598561
KEYWORDS HTG
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 67730)
Lin.X., Kaul,S., Rounsley,S.D., Shea,T.P., Beilto,M.-I., Town,C.D.,
Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,K.E.,
Feildblum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Rongling,C.M.,
Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., Vanhaken,S.E., Umayam,L.,
Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
Nieman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 67730)
AUTHORS Lin.X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Dec 17, 1999 this sequence version replaced gi:4415928.
The sequence and annotation of chromosome 2 were removed from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(http://www.tigr.org/tdb/at/at.html).

```

Genes were identified by a combination of three methods: Gene prediction programs including GRLI (<http://artur.epm.ornl.gov/xgrali>), GeneFinder (Phil Green, University of Washington), GENSCAN (Chris Burge, <http://genomic.stanford.edu/GENSCAN.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curation based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/Mashu/ABI consortium for sequencing BAC clones F6P23, F506, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jintang Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalek, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: atetigr.org.

| FEATURES | Location/Qualifiers |
|--------------|---|
| source | 1..67730 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="II" |
| misc_feature | 1..>67730 /note="Sequence from clone F13A10" complement(join(<1047..1322,1409..1518,1595..1997,2074..2431,2503..2828,2916..3029)) /gene="At2g46480" complement(<1047..>3029) /gene="At2g46480" /note="F13A10.1; predicted by genscan" complement(join(1047..1322,1409..1518,1595..1997,2074..2431,2503..2828,2916..3029)) /gene="At2g46480" /note="hypothetical protein" /codon_start=1 /protein_id="AAD20159.1" /db_xref="GI:4415928" |
| CDs | 1..67730 /translation="MTDACLKNGEDMKVPRFGHTGKAGPDRPMLEHRSIOEK RLERANELMNDSDQKLETAARSRVSDAPRGNTITKNEKRRKSEEDMLRLMD QIARVYSGIAKFTNNLADHOELETDLKLANEESITDIDQRLVDSIRDMGQILA RAHQDYECKLVTKRLAMQLQVEDELENQYITFLQALSAALPDALHCLMRLNTL YHLLPLPMRFPRENTLENPKLYHYALFEDNVLASVYVNSVYVNAADOPSRYHEVLY TDKLNFGAMSWFLLNPGEATIHVORFEDFWLNSYSYVLSLSAAMKFEYETA RSSEVSGSENLKRYRYKMSMLNHEFYIPRIEPLKEKLEFVDDVYVOKDITPLMS IDLRKVENENDRPKCGAYAGMNFQLEKKRNKINETETHEFMQNLNENLTINKTGILP PGLITFNLTQPLQKRWHLGLGIDKQIDVAKIERSAVHIYHNMKPMVEMGSKIQP YWKYTNDEHYITFCRLF" |
| mRNA | 4060..4696 /gene="At2g46490" |
| gene | 4060..4696 /gene="At2g46490" |
| CDs | 4100..4504 /note="F13A10.2" 4100..4504 /gene="At2g46490" |

[illegible]

```

177 hrVala1atrpSerSerAlaproglyIleSerSerAnlySerSerThr 193
||||| : : : : : |||||
9842 CATTAAGATGTCGGGTTCTCGCGAGATTATGCCAAGAAATCAAGATTC 9793
194 LeuSerSerThiHisGlnGlyLeuAlaTyrGlyPheGluLeuSerTyr 210
| : : : : : | : : : : : |
9792 TCACAGTATGTCACAGATTAAGACCTTTGG.....CTGAATGGGA 9752
210 sarGAsnLeuLeuCyAsGAsnCyAsPArg 220
||| : : : : : |||
9751 T...TCTCCAGTGTGAGGATTTGAGAGG 9724

seq_name: gb_pl:AF237569
seq_documentation_block:
LOCUS AF237569 3300 bp DNA linear PLN 08-MAY-2000
DEFINITION Oryza sativa receptor-like protein kinase (RLG14) gene, complete
ACCESSION AF237569
VERSION AF237569.1 GI:7716486
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehartoideae; Oryzaceae; Oryza.
1 (bases 1 to 3300)
Yun,C.-H., Lee,G.-R. and Kim,H.-I.
Molecular analysis of receptor-like protein kinases in rice
Unpublished
2 (bases 1 to 3300)
Yun,C.-H., Lee,G.-R. and Kim,H.-I.
Direct Submission
Submitted (22-FEB-2000) CytoGenetics, National Institute of
Agricultural Science and Technology, Seodundong 249, Suwon 441-707,
Korea

FEATURES
Source
1..3300
Location/Qualifiers
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
577..2791
/gene="RLG14"
Join(577..1295,1411..1568,1644..2791)
/codon_start=1
/product="receptor-like protein kinase"
/protein_id="AA068399.1"
/db_xref="GI:7716487"
/translation="MAHGVFHCSTAILRVISVCLLAVIVPDAGRRHRRGCPPT
CGNLSSVSPPRRRGGPSECVSTELNCDADATQIDNGITYYTGIVNTSTLIIV
DADISDSSCPLEPRMNSLEPRVOVLMTRIASFVSCSEIKSSWIMVPACLSTS
RSFVYVLTGKSAIQLSPSCGLYAMTPUGSDVAPVNESLSYADVVSXMRGF
RVQPEPRYRGTSFQCPMDISIPGKTNWGTILSADTNFLCLIPVQIPSRYRMLGI
MLYVIALQIWKIATCFRLPLAVLYFLAKHKTWRTIDAVEKRLRQOMGIP
TRYAVTDITATIGHERKLGOGGYGVYGVLLPGLHAAVKMLNNSNCGNEEFIS
VATIGMHVNVAVIANGVSCSEMRALVYEMRPGSLDXYIYLSPEWMDKINTEL
GIARGINYLHQGCDMILHFDIKPHNILLDXNVVYVADFGALKYPRKSTVPDRAL
RGIVGMAPEVRSFVSSKSDVTSFGMLLEMGAGRRNDVPPNADSSKSKAYTPSW
VYDGLIADQGFDEISVANNHELEKICLIGLACIKMKSHDRPTSEALEMEGGVDA
LQVPPRPFGDGDIGNGMPPQVMDSYFHSKLTITISEGDIATLAAS"

BASE COUNT 861 a 681 c 771 g 985 t 2 others
ORIGIN

alignment_scores:
Quality: 162.00 Length: 197
Ratio: 1.361 Gaps: 9
Percent Similarity: 60.406 Percent Identity: 23.858

alignment_block:

```

```

US-09-805-919-1 x AF237569
Align seg 1/1 to: AF237569 from: 1 to: 3300
29 CysValProSerSerCysGlyAspIle...GlnIleLysPheProPheAr 44
||| : : : : : |||
694 TGCCCTCCCTTCACCTGTGTATCTTAGACAGCGGTGTGCTTCCTTCG 743
44 gleuArgThrAspProGluHisCysGlyArgArgGlyTyrGluLeuAsp 61
||| : : : : : |||
744 CCGCGCAGGTGACCATCTGAGTGGCGAGTCTGTGCGACAGCTCAT 793
61 ysoLAsnAsnGlnThrValPheAsnTyrLysSerArgIlePheAspVal 77
||| : : : : : |||
794 GTGCCAGTACCAAGCTCAATTCAGATTGACACGCAACATCTATG 843
78 GlnGluIleAsnTyrArgSerTyrSerIleArgLeuLeuAspProGly 94
||||| : : : : : |||||
844 ACGGGCATCACTACCTGATCTTCTTGTGATGCTGTGATGCTGACAT 893
94 uAsnAspGlnArgGluAsnCysThrValPheProAsnHisArgLAsn 111
||||| : : : : : |||||
894 CTGGATTCACCTAGACAGTGGCCCTCT.....CTCGAT 928
111 yAspAlaMetThrSerGlnIlePheGlu...TyrValArgValAsn 125
929 GGATTTCCCTTCGAGACCCAGCTTCAGGTTCTAGTATGACATGATA... 975
126 AsnAspIleAsnTyrValAsnCysLeuAlaProIleAsnSerSer.... 140
976 .....GCTAGCTTGTCTATGTGTTCGACGAGAAATAGATATGCTG 1019
141 ...GlnTyrIleProThrSerPheCysSerLysAsnSerThrGlyPhe 156
||||| : : : : : |||||
1020 GATTATGTCATGACGCTGCTGCT...TGCTACGACACGACGACGCTCG 1066
156 eTyrLeuValIleArgGlu.....IleGlnAlaSerAspLeu 169
||||| : : : : : |||||
1067 TTATGTGTACTGCGCCAAATGCTCTTATTCAGGCTTGAGCT 1116
170 AlaGlyGlyCysArgValGluThrValAlaTrp.....SerSerAl 183
||||| : : : : : |||||
1117 TCTTGAGTACCTAGCCATGACCTTGGGTGCGGACAGTATGAC 1166
183 aPro.....GlyIleSerSerAsnLysSerSerThrLeuSerTyr 197
||||| : : : : : |||||
1167 ACCAGTAATGAAGCTTAAGCTTAAGCTATGACAGCTAATAACTA 1216
197 hrHisGlnGlyLeuAlaTyrGlyPheGluLeuSerTyrLys 210
||||| : : : : : |||||
1217 TGAGGAAGATTCAGAGTCCAAATTTCCTTCAGATATCGT 1257

seq_name: gb_pl:AP002867
seq_documentation_block:
LOCUS AP002867 144322 bp DNA linear PLN 26-JAN-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:p0463F06.
ACCESSION AP002867
VERSION AP002867.1 GI:10179052
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (cultivar:Nipponbare) DNA, clone:p0463F06.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehartoideae; Oryzaceae; Oryza.
1 (bases 1 to 144322)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(ga3) genomic DNA, chromosome 1, PAC
clone:p0463F06
Published Only in Database (2000) In press
REFERENCE
2 (bases 1 to 144322)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
TITLE

```

JOURNAL

Submitted (13-SEP-2000) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
Genes were predicted from the integrated results of the following: GENSCAN1.0, BLAST2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI Nonredundant Protein database, nr

(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is classified based on the 'protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.
The orientation of the sequence is from SP6 to T7 of the PAC clone. Detailed information on overlap and assembly quality together with annotation of this entry is available at
http://rgrp.dna.affrc.go.jp/genomeseq.html.

FEATURES

SOURCE

1. 144322
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0463F06"
join(3115..3947,5188..6259)
/gene="P0463F06.1"
join(3116..3947,5188..6259)
/gene="P0463F06.1"
/note="probably inactive due to frameshift in CDS pseudogene, receptor serine/threonine kinase"
/codon_start=1
/pseudo
complement(join(8047..9173,9246..9430,9518..10293))
/gene="P0463F06.2"
complement(join(8047..9173,9246..9430,9518..10293))
/gene="P0463F06.2"
/codon_start=1
/product="putative receptor kinase"
/protein_id="BAB17115.1"
/db_xref="GI:11034591"
/translation="MAVAGAFRRSTSTALQALTSVLAALQVGGQMGPEEAQTA QCPFSWGLKNVSPPRRHGDDPDGYSKSHELVCSDTATIRINNATYVRIIHW LFTVVDLKNNSSCPRLRQHPHERISMAFSDGVYELAPFNMAAVNC SKETIRNVRKPAVACSTSHSEFYVLGSESEVYVESLEPGYIGMTPLDDGASI NASISYADVKKYMRKFGAIGISSSGTYHDPFACIAESVIFROMVPSDSITYKNR LLATLLAYGVKOCIMAPYKLPVRAVEPAITGCPKMSILTWINVCRLAPLYVL FLSEKIRKARITIDAVKRLQEMLSPMKYGITDIALTSFRRDRLGSGSYGYKQ VLLPGNTHAVAKMLNNGNSNGEELISEVSTIGRIHNVAVRLVGFCSKEMRALVE YMPGSLDKYIFSSDKFSRDKLNEIALGIRGINIHQGCQDQILHEDIKFNHLLD KNFVPKADGGLARLYERDKSFYVSAAGTAVYIAEMISRSFYISKSSEYFGM LLEMGAGRRADPNAENSSOAYPSRYRQLTROETGITAADHEKELCIYGL WCIOMRSCDRPMKSEVIEMLGGVDQLQIPRRFFDDDIYIAMESTYLSSELELAI SEEDRESITELN"
join(11349..11391,11477..12273,12952..14085)
/gene="P0463F06.3"
join(11349..11391,11477..12273,12952..14085)
/gene="P0463F06.3"
/codon_start=1
/product="putative receptor kinase"
/protein_id="BAB17116.1"
/db_xref="GI:11034592"
/translation="WTRTIPILQRLQIYLLSLNVEYSTRKADDEGFLKSCSSQ R CSKHGPDIREPFRSLTSPCGMPGMQLSCSGDITLDHPVLGSKVAYIYRYVIN AILLVESSNCPLOKLASTNVSTAVYEPKDKDAITVCGSKSIATNOGIVGPSCL NLSHSGSQLWLYVPPRYTMSALPPDCTVAVAKGIPYINYNDRGNPNQGLD:SNLKE

ANKVINGEIAFTWHLNKTITNACQCEENMGHCGFSSQNGAFCHQOGLILHFFPK GIQIILAHITLIMEDQTNFYRTYATLSAATLYLLMAGTITYSTIRNNEHLTK VNEILGTGSKREPTRESEVEKRTIPRSEKTHGQGVNGEIPNCPVAMWLN SRDQGEPIFNEVITGRIHANVRLILGCGSCGRTIYEZMPSDEKRTISQGN VSEEFPLDMNDISIGIRKHEYLHOGCNOILTFDIPQILDIYSFRIIDFGL AKICARDQSLVITFARGTMGYIAPELSRNGVISTYSDVSFQMLVLEWGSRL DPQIDNOHEVYFLEWYERKVFQONLITQENTQOEKVRRLAIVALMCIOWMKNR PSTQOVNMTGRIQDLOMPRPFPVSSAESIP"
complement(join(14626..14732,14821..15151))
/gene="P0463F06.4"
complement(join(14626..14732,14821..15151))
/gene="P0463F06.4"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB17117.1"
/db_xref="GI:11034593"
/translation="MGQLDGSRSLTFMEDCMQEGGIRAIAPSTIEHPVIRRRPV AELIKKAWIRDRISGLAQVAIREYKIKSLNONGKCPSPQDYCKMKDSSGYSK SATRALFLDAGQRIDYVVAHILDAIFVITRVLITFLA"
16173..17087
/gene="P0463F06.5"
16173..17087
/gene="P0463F06.5"
/codon_start=1
/product="putative receptor kinase"
/protein_id="BAB17118.1"
/db_xref="GI:11034594"
/translation="MATAVLYLSKRYDEEHLKYENFLRYTSGKPYRFSQVKRI TRCFKEQGGGGSVNIIVLLGFCSEGRHLIYEFMNESELEYIFHPNPSQEL LARKMIDILGIARMEYHOGCNRNLIHPIKRNHLLDYNFLKISDGLKCLA RDSIVITLRAGTMGYIAPELSRNGEISYKSDVSPGMVLEWGSRRNSPSE DQNEVYFLEWYERKVFQDQVLSGEMPEDEKTKROMALVALMCIOWMKNRSMK VNMKLRQNLQIVPPRPVYSIESHAVP"
complement(join(17879..18106,18186..18372,18654..18962, 19113..19291,19329..19461,19516..19624,20367..20631))
/gene="P0463F06.6"
complement(join(17879..18106,18186..18372,18654..18962, 19113..19291,19329..19461,19516..19624,20367..20631))
/gene="P0463F06.6"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB17119.1"
/db_xref="GI:11034595"
/translation="MGCGQEOHRSRSCSRSPASSCSWSTSSMATGAVGEYRAG LVDERQPYTELARLHCRPSIPSPPELVVPPSPVYGGELARGEDVGLLSAQ GCGGRBRRPPTGPMMAAGCAAYVDPIERRELVGCMRPRPCHAGCKRACY EADEDEDAKSRGRPLGKILIAVDVODITLDMQJGSIYGKTYAFIRDFPRLQ NQNHNDIOWMTWTISSVRLQVANAANSNSRYLAPSAALYARFLRPLPAADPNA VVNPAAVLAASDLARIGVSAKKAAYLHDLGRPAAGELSSAAVAADDEAPAPSRQ PTLSPAPAYENKESREERRRRGEERVMTVPADMMGPRGSHADSAYTDKGMRTTEG PRLNLACRTSIPFPPLATNPSPAIPFTPGHLCFQDVALSRAVNAACPGVAP SPCLVAPPEQIOAAPPOD"
join(22482..22821,23165..23209,24203..24528)
/gene="P0463F06.7"
join(22482..22821,23165..23209,24203..24528)
/gene="P0463F06.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB17120.1"
/db_xref="GI:11034596"
/translation="MPARCSHRPRSSAGRADHMOULSPPPYATDTTITATIMDE QOETKRTSKCMPRIIPPARLTSNRQIYIPPSARLYLWESSWSPISNGFSISK SHLWIPQSKQGIELICTGKINPRVAVAPCAGARARAPAAARGYSGRGATAYCS RCPRAVATGSAVPRARLARLPWPLMTSSHQCAPOSLTFENAVAMVNCRTTGP ROIGDCDSYGLSPYVSD"
join(29524..29612,29887..29965,30113..30169,30207..30533)
/gene="P0463F06.8"
join(29524..29612,29887..29965,30113..30169,30207..30533)
/gene="P0463F06.8"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB17121.1"
/db_xref="GI:11034597"

